Krishna's

Elements of STATISTICS

As per U.P. UNIFIED Syllabus (*w.e.f.* 2012–2013)
(*for B.C.A. IIIrd Sem. Students of all Colleges Affiliated to Universities in Uttar Pradesh*)

By

J.P. Chauhan
M.Sc. (*Mathematics*)
C.C.S. University, Meerut

Sharad Kumar
M.Sc., M.Phil., (N.E.T.)
Dep't. of Mathematics
B.S.A. (P.G.) College, Mathura (U.P.)

KRISHNA Prakashan Media (P) Ltd.
KRISHNA HOUSE, 11, Shivaji Road, Meerut-250 001 (U.P.), India

Jai Shri Radhey Shyam

Dedicated

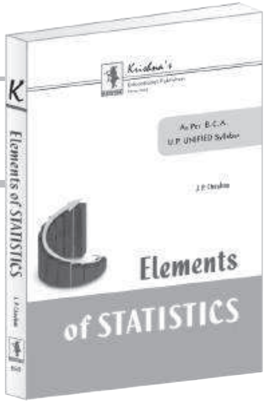
to

Lord

Krishna

Author & Publishers

ABOUT THE BOOK



Introduction

Elements of Statistics: It has been written to fulfil the needs of Ist Sem. B.C.A. Students. It is needless to say that the Matrix, Sets, Arithmetics. Detailed chapters of Differentiation and Integration, A.P., G.P. and H.P. gives the better outline to acquire Mathematical knowledge.

Features

Accessibility: Each chapter of the book has been written in a simple & lucid style so that it is easily accessible. The basic idea of each topic is carefully and clearly explained.

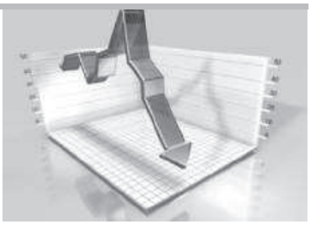
Prerequisites: Prerequisite topics of each chapter are listed in the contents of the book and a basic understanding of those topics will suffice for the study of that particular chapter.

Exercise Sets: Each chapter includes a comprehensive exercise set containing routine, intermediate and challenging exercises so that the learners may have sufficient practice in the subject. Solutions to all exercises are given at the end of each chapter.

Goals

The goal of this book is to present a wide variety of interesting applications of B.C.A. Statistics in a flexible and accessible format. The book entitled **Elements of Statistics for B.C.A.** has been written to meet requirements of B.C.A. course of all Indian Universities on this subject. It covers topics from Sets, Percentage, A.P., G.P., H.P., Differentiation and Integration, Matrices. This book contains illustrative examples, theorems with proofs and comprehensive exercises. Throughout the book, the subject matter has been dealt in such a simple, logical and systematic way that students will not find any difficulty to understand it.

PREFACE



It gives me pleasure to place the book **'Elements of Statistics for B.C.A.'. (All Universities in U.P.)** which has been completely revised and made upto date in the light of suggestions received from the students and the learned teachers of various universities.

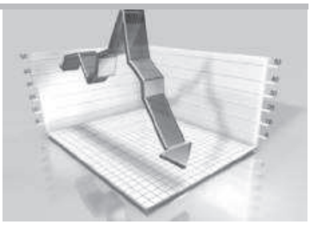
The following are the main features of third edition:

- More than 150 new problems and examples have been added.
- Includes recent elegant and elementary proof.
- All chapters are arranged according to the latest **U.P. Unified Syllabus** implemented from 2012 – 2013, keeping in view the requirement of the students of B.C.A. IIIrd semester of all Universities in Uttar Pradesh (U.P.).
- Errors and omissions have been corrected.
- Up-to-date yearwise references from various universities examination papers have been given throughout the book.
- Each step is made clear by giving reasons with clarification.

We hope that the book in its present form will prove beneficial to the students. We shall also forward to receive your suggestions and comments bout the book so that it could be further improved.

— J.P. Chauhan

ACKNOWLEDGEMENT



This book has been written in simple and lucid style for better understanding and appreciation of study. It has been compiled systematically to enable sure one to evaluate detail of the book.

Needless to say, I am deeply indebted to the publisher **Shri Satyendra Rastogi** and **Shri Sugam Rastogi** for providing continuous support and enthusiasm.

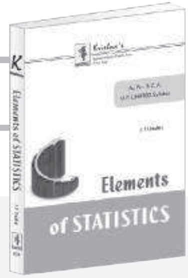
This acknowledgement would fall short of completion if I fail to mention the **entire team of Krishna Prakashan.** I would like to thank the contributors for their creativity and especially their excellent work.

Finally, we shall be very grateful to all instructors and students who send us their helpful suggestions.

Meerut (U.P.)

—J.P. Chauhan

SYLLABUS



Elements of Statistics

U.P. UNIFIED (*w.e.f.* 2012-2013)

B.C.A. S. 205 **B.C.A. IIIrd Semester** **3 1 0 4**

Unit-I

Population, Sample and Data Condensation: Definition and scope of statistics, concept of population and simple with Illustration, Raw data, attributes and variables, classification, frequency distribution, Cumulative frequency distribution.

Unit-II

Measures of Central Tendeny: Concept of central Tendency, requirements of a good measures of central tendency, Arithmetic mean, Median, Mode, Harmonic Mean, Geometric mean for grouped and ungrouped data.

Unit-III

Measures of Dispersion: Concept of dispersion, Absolute and relative measure of dispersion, range variance, Standard deviation, Coefficient of variation.

Unit-IV

Permutations and Combinations: Permutations of 'n' dissimilar objects taken 'r' at a time (with or without repetitions). $^nP^r=n!/(n-r)$!(without proof). Combinations of 'r' objects taken from 'n' objects. $^nC^r=n!/(r!(n-r)!)$ (without proof) . Simple examples, Applications.

Unit-V

Sample Space, Events and Probability: Experiments and random experiments, Ideas of deterministic and non-deterministic experiments; Definition of sample space, discrete sample space, events; Types of events, Union and intersections of two or more events, mutually exclusive events, Complementary event, Exhaustive event; Simple examples.

Classical definition of probability, Addition theorem of probability without Proof (upto three events are expected). Definition of conditional probability Definition of independence of two events, simple numerical problems.

Unit-VI

Statistical Quality Control: Introduction, control limits, specification limits, tolerance limits, process and product control; Control charts for X and R; Control charts for number of defective {n-p chart},control charts for number of defects {c - chart}.

BRIEF CONTENTS

Dedication.. (iii)
About the Book...(iv)
Preface ....(v)
Acknowledgement ....(vi)
Syllabus (B.C.A. IIIrd Sem.)..(vii)
Brief Contents ....(xii)
Detailed Contents..(ix–xii)

Chapters in the Book

Unit–I

Chapter 1: Statistics Concept, Significance and Data.................................(01–22)
Chapter 2: Classification and Tabulation of Data..(23–48)
Chapter 3: Frequency Distribution and Graphical Representation................(49–96)

Unit–II

Chapter 4: Measure of Central Tendency..(97–166)

Unit–III

Chapter 5: Measure of Dispersion..(167–218)

Unit–IV

Chapter 6: Permutation and Combinations...(219–262)

Unit–V

Chapter 7: Probability...(263–316)

Unit–VI

Chapter 8: Statistical Quality Control Methods.......................................(317–352)

DETAILED CONTENTS

Chapter-1: Statistics Concept, Significance and Data..........(01-22)

1.1 Definition of Statistics 01
  1.1.1 Statistics in Plural Sense 01
  1.1.2 Statistics in Singular Sense 03
  1.1.3 Distinction between the Two Senses of Definitions 04
1.2 Scope of Statistics 04
  1.2.1 Statistics and State 04
  1.2.2 Statistics and Economics 04
  1.2.3 Statistics and Economic Planning 05
  1.2.4 Statistics and Business 05
  1.2.5 Statistics and Psychology and Education 05
  1.2.6 Statistics and Medical, Biological and Agricultural Sciences 06
1.3 Limitation of Statistics 06
1.4 Distrust and Misuse of Statistics 07
  1.4.1 Reasons for Distrust in Statistics 07
1.5 Nature of Statistics (Science or Art) 08
  1.5.1 Statistics as Science 08
  1.5.2 Statistics as an Art 08
  1.5.3 Statistics is both Science and Art 08
1.6 Functions of Statistics 09
1.7 Statistical Units 10
1.8 Collection of Data 11
  1.8.1 Difference between Primary and Secondary Data 11
  1.8.2 Methods of Collecting Primary Data 12
1.9 Framing a Questionnaire or a Schedule 15
1.10 Collection of Secondary Data 16
1.11 Limitations of Secondary Data 17

1.12 Precautions in Accepting the Secondary Data *17*
1.13 Advantages of Secondary Data *18*
Solved Examples 18
Problem Set 21

Chapter 2: Classification and Tabulation of Data..................(23-48)

2.1 Classification *23*
  2.1.1 Objects of Classification *23*
  2.1.2 Types of Classification *24*
2.2 Construction of Frequency Distribution *26*
  2.2.1 Steps in the Construction of Frequency Distribution *30*
2.3 Cumulative Frequency Distribution *30*
  2.3.1 'Less than' Cumulative Frequency Distribution *30*
  2.3.2 'More than' Cumulative Frequency Distribution *31*
2.4 Relative Frequency Distribution *31*
Solved Examples 32
2.5 Three Forms of Data *38*
Problem Set-1 39
Answers 40
2.6 Tabular Presentation *41*
  2.6.1 Table and Tabulation *41*
  2.6.2 Advantages of Tabulation *41*
  2.6.3 Preparing a Table 41
  2.6.4 Requirements of a Good Table *43*
  2.6.5 Type of Tables *43*
2.7 Difference between Classification and Tabulation *47*
Problem Set-2 48

Chapter-3: Frequency Distribution and Graphical Representation...(49-96)

3.1 Display Data *49*
  3.1.1 Diagrammatic Presentation *49*
Solved Examples 50
  3.1.2 Graphical Presentation *62*
3.2 Graphs of the Time Series *77*
Problem Set 93

Chapter-4: Measure of Central Tendency............................(97-166)

4.1 Averages *97*
  4.1.1 Characteristics of a Good Average *97*
4.2 Measures of Central Tendency *98*

4.2.1 Arithmetic Mean *98*
Solved Examples 99
4.2.2 Median *131*
4.2.3 Mode *140*
Problem Set 160
Answers 166

Chapter-5: Measures of Dispersion.................................(167-218)

5.1 Dispersion *167*
5.1.1 Objective of Measuring Variability or Dispersion *167*
5.1.2 Properties of a Good Measures of Dispersion *168*
5.1.3 Two Types of Measures of Dispersion *168*
5.2 Important Measures of Dispersion *168*
Solved Examples 169
List of Formulae 214
Problem Set 216
Answers 218

Chapter-6: Permutations of Combinations........................(219-262)

6.1 Factorial Notation *219*
Solved Examples 220
Exercise 223
Answers 224
6.2 Permutation *224*
Exercise 231
Answers 231
6.2.1 Permutation with Repetition *232*
6.2.2 Circular Permutations *242*
Exercise 244
Answers 245
6.3 Combinations *246*
6.4 Difference between a Permutation and a Combination *246*
Exercise 260
Answers 262

Chapter-7: Probability..(263-316)

7.1 Introduction *263*
7.1.1 Classical Approach to Probability *263*
7.1.2 Classical Definition of Probability *265*
Solved Examples 266
7.1.3 Theorem of Total Probability *274*
7.1.4 Compound Events *275*

7.1.5 Theorem of Compound Probability or Multiplicative Law of Probability *275*

Problem Set 293

Answers 294

7.2 Set Theoretic Concepts *295*

7.2.1 Computation of the Probability of an Event by the Axiomatic Approach *302*

7.3 Conditional Probability 304

7.4 Baye's Theorem 307

Problem Set 315

Answers 316

Chapter-8: Statistical Quality Control Methods..................(317-352)

8.1 Quality *317*

8.2 Method for the Measured the Quality *317*

8.3 Quality Control *318*

8.4 Statistical Quality Control (S.Q.C.) *318*

8.4.1 Frequency Distribution *318*

8.4.2 Control Chart *319*

8.4.3 Acceptance Sampling *319*

8.4.4 Analysis of Data *319*

3.5 Advantages of Statistical Quality Control *319*

3.6 Reasons for Variations in the Quality of a Product *320*

8.6.1 Chance or Random Cause or Non–Assignable Causes *320*

8.6.2 Assignable Cause or Preventable Causes *320*

8.7 Techniques of Statistical Quality Control *321*

8.7.1 Process Control *321*

8.7.2 Product Control *321*

8.7.3 Control Chart *322*

8.7.4 Control Limits *322*

8.8 Object of Control Charts *323*

8.9 Type of Control Charts *324*

8.9.1 Construction of Control Charts for Variables *324*

8.10 Control Charts for Attributes *326*

Solved Examples 329

8.11 Control Charts for Variables Versus Control Charts for Attributes *338*

8.12 Advantage of Statistical Quality Control *339*

8.13 Limitation of Statistical Quality Control *339*

8.14 Control Chart for Number of Defects per Unit (C-Chart) *340*

Exercise 347

Answers 350

❑❑❑

Chapter

1

Statistics Concept, Significance and Data

1.1 Definition of Statistics

[B.C.A. (I.G.N.O.U.) 2011, 2007; B.C.A. (Lucknow) 2008, 2006; B.C.A. (Rohilkhand) 2008, 2004; B.C.A. (Agra) 2007; B.C.A. (Meerut) 2000]

There have been many definitions of the term 'Statistics' which can be broadly classified in two senses:

1. Statistics in Plural Sense
2. Statistics in Singular Sense

1.1.1 Statistics in Plural Sense

As a plural noun statistics is used for denoting numerical and quantitative information. Thus, in plural sense, it means the same thing as data. For example, statistics of scores in a cricket match, trade related statistics, labour statistics, census statistics, godown statistics and so on.

All the quantitative observations are not statistics but those quantitative observations which satisfy certain conditions are called 'Statistics'. Also, economist **Horace Secrist**, defined 'Statistics' as:

"By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in a relation to each other".

According to this definition of statistics, the quantitative expression should satisfy the following conditions:

(i) **Statistics are a Collection of Facts:** Single or isolated numerical expressions are not statistics as they can't be compared and put in relation. For example, Rohan scored 127 runs in a match, is not statistics but 10 batsmen scored 32, 57, 10, 0, 98, 15, 2, 0, 66 and 7 runs respectively are statistics. Hence, a set or a collection of observation forms statistics.

(ii) **Statistics are Collected in a Suitable Order:** Statistics should be collected with a suitable plan of data collection. Statistics collected without any system or order, will result in failed decisions.

(iii) **Statistics are Collected for a Predefined Purpose:** The statistics should be collected for a specific purpose. For example, while collecting statistics of trading, we must decide beforehand the regions, commodities and the periods for which they are needed.

(iv) **Statistics are Expressed Numerically:** Quantitative expression are not statistics unless are expressed in numerical forms. Statements like 'India's population is increasing', 'prices of petrol will go high', 'Ram is good at Economics' are not statistical statements but statements like 'There is 63% increase in food items', 'Average poverty ratios in India is 26' are statistical statements.

(v) **Statistics are Affected to a Marked Extent by Multiplicity of Causes:** Statistics or data are affected by a number of forces operating together *i.e.*, they are dependent on a number of phenomena. For example, statistics of production of an agricultural commodity are affected by rainfall, soil quality, manuring process and method of cultivation etc. Similarly, it same for other sectors also.

(vi) **Statistics are Placed in Relation to Each Other:** Statistics are collected for making comparisons between related phenomena over a specific region or period. For example, the population figures of India, UK, USA, China etc. in 2011 census are comparable as they belong to same census year. Thus, for a valid comparison, statistics are required to be put in relation to each other.

(vii) **Statistics are Enumerated or Estimated According to Reasonable Standards of Accuracy:** Enumeration means a precise and accurate numerical statement. These statements can be obtained by two ways either by actual counting or measurement or by estimate. For example, the number of students in class can be counted with perfect accuracy but in situation, such as in a rally, one has to estimate the number of people. Similarly, exact measurements on heights, weights, distances etc. are not possible and one has to make estimates with reasonable standards of accuracy.

1.1.2 Statistics in Singular Sense

Whenever a huge amount of data is collected, there arise a need to organize, present, analyse and interpret them in a systematic order. Statistics, in a singular sense is a branch of science which deals with scientific methods of collection, organization, presentation analysis and interpretation. According to **Croxton** and **Cowden**, "Statistics may be defined as a science of collection, presentation analysis, and interpretation of numerical data".

Thus, we have the following five stages in a statistical investigation:

(i) **Collection of Data:** With a specific and well defined purpose is the first step in a statistical investigation. If the collected data are faulty, the conclusions drawn are bound to be fallacious. Hence, data should be collected with a maximum core as they form the foundation of statistical analysis.

(ii) **Presentation of Data:** Collected and organised data are presented in a systematic order to facilitate statistical analysis. The organised data are presented in the form of tables, graphs and diagrams to allow better understanding and comparison.

(iii) **Organisation of Data:** is also an important step in statistics as deriving any conclusions from the observations only is difficult. So, its purpose is to provide a methodology for classification and description in a summary form. Editing, classification and tabulation are three step in organisation of data.

(iv) **Analysis of Data:** Statistics plays two important roles in analysis of data:

(i) Descriptive statistics and (ii) Inferential statistics.

Difference between these two are:

Descriptive Statistics	Inferential Statistics
It is concerned with describing and summarising the numerical properties of data. Its methodology includes classification, tabulation, graphical presentation and calculation of certain indicators of data such as mean range, standard deviation etc. which summarise certain important features of data	It is concerned with certain derivation of scientific inference about generalisation of results from the study of a few particular cases. Technically speaking, the methods of inferential statistics helps in generalising the results of a sample to the entire population from which the sample is drawn.

The nature of inference is that we make general statements from the study of a few cases. In such case, generalisation can't be made with certainty. Thus, language of probability is often used in statistics conclusions.

1.1.3 Distinction between the Two Senses of Definitions

Statistics as Data	Statistics as Method
1. It is plural.	1. It is singular.
2. It refers to a sense of data.	2. It refers to any statistical method.
3. It may be of primary or secondary nature.	3. It is of scientific nature.
4. It is in the form of raw materials.	4. It is in the form of tools applied for processing the materials.

1.2 Scope of Statistics

[B.C.A. (I.G.N.O.U.) 2010; B.C.A. (Lucknow) 2010, 2008, 2004; B.C.A. (Rohtak) 2009; B.C.A. (Meerut) 2006, 2004]

Scope of statistics is indeed very fast. It covers all fields of study where in quantitative appear and need analysis. It is essential from the state of planning to the stage of conclusions. It applies to all sciences and applications, such as: agricultural, engineering, medical, physical, economics, business, commerce, industry etc.

1.2.1 Statistics and State

As per reports of collection of data statistics was developed as a state craft for maintaining economic and military strength and for formulating different policies. Gradually, the sphere of data collection to other fields such as sales, health, taxes, import, export, production etc. Now, the functions of a state have increased due to increase in the concept of welfare state. For *e.g.* Population.

1.2.2 Statistics and Economics

[B.C.A. (A.M.U.) 2008; B.C.A. (Kashi Vidhyapeeth) 2007]

Statistics as 'data' and 'methodology' have proved immensely useful in solving a variety of economic problems related to wages, prices, analysis of time series, income and demand analysis. The development of various economic theories own greatly to statistical methods. *e.g.,* 'Engel's law of family expenditure', 'Mathusian theory of population'. The impact of mathematics and statistics has led to the development of new disciplines like 'Econometrics' and 'Economic statistics'.

1.2.3 Statistics and Economic Planning

[B.C.A. (Kashi Vidhyapeeth) 2011; B.C.A. (A.M.U.) 2011, 2009; B.C.A. (Lucknow) 2010, 2006]

Statistics is the most important tool in economic planning which is essential for proper utilisation of natural resources. Drawing plan, execution and review is based on statistical analysis. Priorities of expenditure of a national budget can be determined through the proper understanding of various economic problems, *viz.*, balance of trade, industrial growth, disparities in income and wealth, agricultural production etc. Time-bound goals are set after listening the priorities of growth in different economic activities . Maximum utilisation of available resources can be made by using econometric models based on statistical analysis.

1.2.4 Statistics and Business

[B.C.A. (A.M.U.) 2011; B.C.A. (Avadh) 2008, 2005]

With growing size and increasing competition, the problems of business enterprises have become complex. Statistics is now considered as an indispensable tool in the analysis of activities in the field of business, commerce and industry. In business, the manufacturer is always interested to estimated the immediate and future demand of his product. The object can be achieved by properly conducted market survey and research which greatly depend on statistical methods. The trends in sales and production can be determined by statistical methods like time-series analysis which are essential for future planning of the phenomena. Similarly, the correlation and regression analysis can be used to judge the effectiveness of various sales promotion measures.

Statistical concepts and methods are also used in controlling the quality of products to the satisfaction of consumer and the producer. Thus, method of quality control and reliability analysis are invaluable in industrial engineering.

1.2.5 Statistics and Psychology and Education

[B.C.A. (Rohilkhand) 2008, 2006]

Statistics has found wide application in psychology and education. It forms the basic of scientific approach to problem in which inductive inference is commonly used. Students of psychology and education cannot remain ignorant of such methods of problem solving in their disciplines. Also, statistical methods are used to measure human ability (for *e.g.*, intelligence, aptitude, personality, interest etc.) by tests. Theory of learning is also based on statistical principles. Application of statistics in psychology and education have led to the development of new discipline called 'Psychometry'.

1.2.6 Statistics and Medical, Biological and Agricultural Sciences

Statistics is also one of the most prominent tool in the field of Medicines, Biology and Agriculture. In Medicine sciences, one is greatly concerned with the results obtained from the use of medicines and drugs. Similarly, the important role of various factors in the growth and development of a plant is important in Biology. In Agriculture, the role of factors *viz.*, manure, rainfall, watering process, seed quality, affecting the production need to be analyzed for good results statistical techniques like 'analysis of variance' and 'Design of experiments' are useful for isolating the role of such factors.

1.3 Limitation of Statistics

[B.C.A. (Kurukshetra) 2011; B.C.A. (Purvanchal) 2009, 2007; B.C.A. (I.G.N.O.U.) 2007; B.C.A. (Kanpur) 2007, 2005; B.C.A. (Meerut) 2004, 2002]

Though, statistics has many application in various important areas but it also has some limitations. Hence, for proper use of statistics, it is also necessary to know the limitations and misuse of statistics. The following are the limitations of statistics:

1. **Statistics can be Misused and Misinterpreted:** The greatest limitation of statistics is that it is likely to be misused. Inadequate and faulty procedure of data collection and inappropriate comparisons may arrive at fallacious conclusions.

   For *e.g.,* when conclusions are based on incomplete information and are drawn by unskilled investigators, then it is said to be the misuse of statistics.

2. **Statistics Deals Only with Quantitative Data:** If the study yields qualitative data which cannot be meaningfully converted to quantitative data, then valid conclusions cannot be drawn using statistical analysis.

3. **Statistics Deal with Aggregates of Items and not with Individual:** Statistics is the study of mass of data and deals with collective information. A single data or individually constituted data is not considered as a statistical analysis. For *e.g.* Akash scored 81 marks out of 100 in Economics, is not a statistical data, whereas, Akash's average marks are 92 is statistical data.

4. **Statistical Laws are True Only on An Average:** Laws of statistics are not universally applicable as the laws of physics, chemistry and mathematics. These may not be true for particular individual. If it is statistically established that annual income of a man is 3 lakhs to 5 lakhs then this statement is true only on an average and not with individual.

5. **Statistics is Only One of the Methods of Studying a Phenomenon:** Statistical methods don't always provide the best possible solution to a given problem. In varying culture and religious situation, it may fail to reveal the underlying facts responsible for the variations in a phenomenon under study. Thus, statistical conclusions are to be supplemented by other variations.

1.4 Distrust and Misuse of Statistics

[B.C.A. (B.H.U.) 2011, 2008]

We have observed that statistical techniques are important tools in analysing problems in different areas of life. But, still there is a considerable distrust in statistics *i.e.*, lack of faith and confidence in statistical methods and statements. It is often said that:

1. Statistics is another form of lying.
2. There are three kinds of lying – lies, damned lies and statistics.
3. Statistics can prove or disprove anything.
4. Statistics reveal what is convincing but conceals the truth.

1.4.1 Reasons for Distrust in Statistics

[B.C.A. (Kanpur) 2008, 2004]

(i) Figures are innocent and convincing due to which unawared people believe them easily.

(ii) Statistical data used for providing arguments may be inaccurate and incomplete and will lead to fallacious conclusions.

(iii) Even if the figures are correct, they may be presented or interpreted in such a manner that the concerned persons get misleaded. *For e.g.* 2 companies X and Y are paying ₹ 30,000 and ₹ 45,000 per month respectively. Now, by this statement company Y seems more interesting. But, it would be misleading until we know the number of employees and the number of working hours etc. Thus, in such cases, we might draw different conclusions, Hence in the view of this statistics, it should be used with utmost carefulness. Thus, it is commonly said that 'statistics neither proves anything nor disproves anything'. It is only a tool, which will be good if used in proper manner otherwise it may lead to bad decisions.

The fault doesn't lie in statistics but this distrust arise because of its misuse on the part of untrained practitioners of statistical methods, ignorant users of statistical results and the dishonest people who use statistics to mislead and confuse rather than to clarify and illuminate.

"Science of statistics is the most useful servant but only of great value to those, who understand its proper use".

For e.g., Hockey is a nice game for physical exercise but wrong conduct of the game and wrong rules & regulations may make it a dangerous game. But, game of Hockey can't be blame for a wrong outcome. Similarly, if the statistical facts are mis-interpreted by the people to serve their ends, it will be unjust to blame the science as such. Instead, people are responsible for these faults.

1.5 Nature of Statistics (Science or Art)

[B.C.A. (Bundelkhand) 2008; B.C.A. (Kanpur) 2007; B.C.A. (Bundelkhand) 2006]

1.5.1 Statistics as Science

Science refers to a systematised body of knowledge. It studies cause and effect relationship and attempts to make generalisation in the form of scientific principles of laws. Like other sciences, statistical methods are also used to answer the question like, how an investigation should be conducted ? In what way the valid and reliable conclusions can be drawn ? How far are these conclusions dependable ? Hence, statistics is a science. The science of statistics is different from physics, chemistry etc. Statistics is called the science of scientific methods. That is why some statisticians do not consider it a pure science but a scientific method. According to **Cowden** and **Croxton**, ''Statistics is not a science, it is a scientific method''. In the words of **Tippet**, ''As science, the statistical method is a part of the general scientific method and is based on the same fundamental ideas and processes''.

1.5.2 Statistic as an Art

An art is an applied knowledge . It refers to the skill of handling facts so to achieve a given objective. It is concerned with the ways and means of presenting and handling data, making inferences logically and drawing relevant conclusions. Statistics is an art which tells us how to use statistical rules and principles to study the problems and finding their solutions. By the application of statistical method a number of complex problems can be understood easily. Collection of statistical (data), it use and utility are itself an art. How the desired results are obtained by using the rules of statistics, which statistical methods are used to solve special type of problems ? How the two facts are compared and by establishing correlation conclusions are drawn ? All these aspects actually relate to art and this is the reason why statistics is an art.

1.5.3 Statistics is both Science and Art

After studying, the science and art aspect of statistics it is said that statistics is both science as well as Art. Statistics is used not only to gain knowledge but also to understand facts and draw important conclusion from it. This type of conclusions prepare a path of the economic and social progress. According to **Tippet**, ''Statistics is both a science and an art. It is a science in that its methods are basically systematic and have general application and an art in that their successful application depends to a considerable degree on the skill and special experience of the statistician and his knowledge of the field of application, *e.g.*–Economics''.

1.6 Functions of Statistics

[B.C.A. (Rohtak) 2010; B.C.A. (Lucknow) 2007]

1. **It Simplifies Complex Data:** Statistical methods like averages, totals, percentages, etc. help in condensing mass of data into a few significant figures so as to make them easily understandable.
2. **It Presents the Facts in a Definite Form:** Statistics deals with the quantitative statements of facts and thus presents them in a precise and definite form. Quantitative statements of facts are always more convincing than vague statements. For example, the statements like 'the birth rate in India is decreasing', 'the prices of commodities are rising', 'Mr. Yogesh has passed the high school examination', etc. do not convey information in a definite form as they do not involve quantitative statements of facts.
3. **It Provides a Technique of Comparison:** Comparison of quantitative facts is also an important function of statistics. For example, area-wise and period-wise comparison of data related to production, sales, export, import, population, etc. are helpful for drawing valid conclusions about economic activities.
4. **It Studies Relationship:** Correlation analysis is used to find the functional relationship between two or more variables. The relationship is very helpful in making estimates and forecasting future trends. For example, the relationship between supply and demand, advertisement and sales, etc. provide us the basic tool for planning these activities.
5. **It Helps in Formulating Policies:** As discussed earlier, inferential statistical methods are helpful in formulating various economic and business policies. Empirical studies of data concerned with imports, exports, wages, production, etc. are helpful in framing suitable policies and plans in the respective areas. Some well established law such as 'Malthusian theory of population' and 'Engel' law of family expenditure' are based on statistical analysis of relevant data.
6. **It Helps in Forecasting:** Statistics is now an indispensable tool in the analysis of activities relating to business, commerce and industry. Such an analysis is useful for determining trends in activities related to these areas, which in turn form the basis of estimation and forecasting about the phenomenon under study.
7. **It Helps to Test and Formulate Theories:** The function of statistics is also to test and formulate policies in different fields and to measure their effects. For example, statistical methods are used in agricultural and biological science in determining the role of various factors of growth and development of the organism or plant under study. In medical sciences, the effects of drugs are recorded and analysed. In all such studies, effects of different factors can be tested by collecting and comparing the relevant data. Properly conducted consumers and market surveys are indispensable tools in formulating precise and definite production policies.

1.7 Statistical Units

1. **In Practical Field of Survey** the definition of location of statistical unit remains a pre-requisite for the collection, interpretation and presentation of the numerical data. A unit stands for the qualitative or quantitative measure in terms of which the data are collected and analysed.

2. **Statistical Units can be Divided into the Following Three Categories:**

   (i) **Unit of enumeration** is that in terms of which objects are measured and collected. Such units may be simple or composite. A simple unit denotes only one attribute or consideration without any qualifying limit attached to it. Suppose at a particular hours and place 4000 passengers travelled over N-railway and 600 over S-railway. Then the comparison of popularity can not be made without knowing the distance travelled in both the cases. If it is found the N-railway passengers covered a distance 200 kms and those of S-railway 125 kms, then to have a correct comparison of the popularity of two trains, the total number of passengers-kilometers in both the cases are obtained, which is

   $$200 \times 4000 = 800{,}000 \text{ for N-railway}$$

   and $$125 \times 6000 = 750{,}00 \text{ for S-railway}$$

   indicate that former one is more popular. The other examples of composite unit, naan-hour, killo-watt-hour, rupee-gram etc.

   (ii) **Unit of analysis and interpretation** is that in terms of which data are analysed, compared and interpreted rates, percentages are used for this purpose.

   (iii) **Unit of presentation** is that in terms of which data are presented. However, ultimately the nature of such units will belong to either of the first two categories only. Because data are presented with respect to time place and condition either after measurement and collection or after its statistical analysis and interpretation.

3. **Characteristics of a Good Statistical Unit**

   (i) Unit should be clear and unambiguous otherwise different people will give different interpretations. For example, the unit of household has to be defined specifically whether it will stand for a family related by blood, or for those living together collectively or having common kitchen.

   (ii) **Unit should be self-explanatory.** The persons for whom survey or the investigation has been conducted should be able to understand and interpret the unit easily.

   (iii) Unit should be rigid and definite without any flexibility in nature.

   (iv) **Unit should be relevant.** For a unit relevant and suitability to the purpose of enquiry is essential. For example, natural income should be measured in national coins. Similarly, for the study of giving conditions of the people, house hold is a suitable unit.

1.8 Collection of Data

Collection of data *is the first step in any statistical investigation of a phenomenon. We know that building cannot be constructed without the right bricks, similarly we cannot come to any conclusion if the collection of data is not proper.*

Data are generally classified into the following two groups:

1. Internal Data 2. External Data

1. **Internal Data:** Internal data *comes from internal sources related with the functioning of an organization or firm* where records regarding purchase, production, sales, profits etc. are kept on a regular basis. Various Government departments like Railways, Communications and Education *etc.*, also generate internal data which are useful for their proper internal functioning. However, the internal data can be either insufficient or inappropriate for the statistical enquiry into a phenomenon. In that situation we need external data.

2. **External Data:** External data *are collected and published by external agencies.* This type of data can be obtained from primary source or secondary source. Thus, the external data can further be classified as: Primary and Secondary data.

   (i) **Primary data** *are original and first hand information.* Data are termed primary when the reference is to data collected for the first time by the investigator. For example, the Meteorological department regularly collects data on various aspects of the weather and climate such as amount of rainfall, humidity, minimum and maximum temperature of a certain place. These constitute primary data. Similarly, the data in a population census obtained by the office of the Registrar General and Census Commissioner are primary in nature.

   (ii) **Secondary data** on the other hand, data are termed secondary when collected from records or data already available. In other words, the secondary data *are one which has already been collected by a source other than the present investigator.* For example, population census data are primary for the office of the Registrar General and Census Commissioner whereas for other organizations or individuals who use such data, they are secondary. Thus, data which are primary in one hand become secondary in the hands of others.

1.8.1 Difference between Primary and Secondary Data

[B.C.A. (Bhopal) 2010, 2006; B.C.A. (Meerut) 2002]

(i) The primary data are absolutely new and are collected a fresh whereas the secondary data are those which were previously collected and published by some previous investigators and are being simply used by the present investigators.

(ii) The primary data are collected according to the plan of the investigator and so it requires much money, labour and time in collecting them, whereas the secondary data is always readily available and can be borrowed easily by the present investigator for his purpose.

(iii) The primary data are collected according to the requirement of the problem under consideration while, the secondary data was collected for another purpose and is moulded according to the present requirement.

1.8.2 Methods of Collecting Primary Data

Primary data are collected for the first time through census or sample surveys. The following are the methods of collecting primary data:

(i) Direct Personal Interviews or Observations
(ii) Indirect Personal Interviews or Observations
(iii) Mailed Questionnaires
(iv) Schedules through Enumerators
(v) Information through Local Sources or Correspondents

(i) Direct Personal Interviews or Observations: In this method the investigator contacts the source of information directly and personally. This method of collecting data is suitable in the following situations:

(a) When the area of enquiry is limited.
(b) When high degree of accuracy is needed in the results.
(c) When the investigation is difficult and is to be conducted in a heterogeneous area.
(d) When the information collected is of confidential nature.

✦ **Merits** **[B.C.A. (Agra) 2003]**

(a) Information collected by this method is accurate and reliable. It can be checked and rechecked and put in the desired form.
(b) Alternative supplementary information can be gathered simultaneously.
(c) Collected information can be made free from bias and prejudices.
(d) This is the most suitable method when the area of investigation is limited.

✦ **Demerits** **[B.C.A. (Agra) 2003]**

(a) Such a method can be adopted only when the enquiry is intensive and localized. It cannot be adopted when the field of enquiry is wide.
(b) The involvement of personal bias and prejudices is obvious.
(c) Investigators should be trained and efficient persons, otherwise the information collected may be unreliable.
(d) It is not scientific.
(e) It requires too much time and labour.

(ii) Indirect Personal Interviews or Observations: In some cases, the informants cannot be contacted directly. In this situation an indirect personal enquiry is conducted to get the desired information. The indirect personal investigation is done through some agencies who have some knowledge of the phenomenon under enquiry. For example, an investigator may collect information about cost of cultivation indirectly from the village Pradhan (headman) who have intimate knowledge of such matters. This method is useful in the following cases :

(a) When the area of investigation is large.

(b) When the information cannot be obtained directly from the informants.

(c) This method is generally used by the government or committees and commissions appointed by the government. They collect information about different problems by interviewing the person concerned.

✦ **Merits**

(a) It is useful when the area of investigation is wide.

(b) It saves labour, time and money.

(c) As the information is collected from knowledgeable persons, it is expected to be useful and reliable.

(d) The enquiry could be more extensive and different aspects of the problems can also be studied properly.

(e) This method is easy.

(f) It is free from the prejudices of the investigator.

✦ **Demerits**

(a) The information collected may be subjective and is subject to the personal bias of the persons or agencies from whom it is collected.

(b) One has to be very careful about the selection of persons from whom the information is collected, because their knowledge and personal attitude towards the problem greatly affect the quality of data. So one has to be very careful in dealing with such situations.

(iii) Mailed Questionnaire Method: In the mailed questionnaire method, a questionnaire in the form of a set of questions is sent by mail to the informants. They are expected to answer the questions and also the additional needed information, whenever needed and mail them back to the investigator. This method should be used in the following cases:

(a) When the area under investigation is wide.

(b) When the informants are educated.

(c) When the informants are expected to leave for far away places.

✦ **Merits**

(a) It is economical and less time consuming.

(b) A larger area can be covered under the investigation as even the people in distant places can be reached without much difficulty.

(c) Collected information may be more reliable.

(d) It is free from the bias of the investigator as the information is collected from the informants themselves.

✦ **Demerits**

(a) This method can be adopted only in cases when the informants are educated persons.

(b) All the informants may not co-operate and the information to be collected may remain incomplete. Sometimes people may not spare time to fill the questionnaire or they may not be willing to return the filled questionnaire. Obviously collected information will not be completed.

(c) The greatest disadvantage of this method is the unwillingness of persons to reply at all or in full due to carelessness or fear of exposure.

(iv) Schedule Through Enumerators: Initially let us make a distinction between a **questionnaire and a schedule.** The questionnaire is a set of questions and the answers to which are recorded by the informant itself, whereas in a **schedule** answers are recorded by the investigator or an enumerator on his behalf. In this method the investigators or enumerators approach the informants with a prepared questionnaire and get the replies to the questions. This method is generally used in censuses and large scale surveys.

In the case of census, investigators visit every member of the source of information in the zones while, in the case of sample survey, they collect information from those members who have been selected in the sample.

✦ **Merits**

(a) This is the only method which is applicable in census and sample survey both.

(b) Subjectivity on the part of the interviewer is less in this method.

(c) This method is useful where the scope and coverage is very large.

(d) Personal contacts increase the reliability of data.

(e) Information from an illiterate person too can be obtained.

✦ **Demerits**

(a) The method is quite costly and time consuming.

(b) Trained and efficient investigators will be needed to get the information.

(c) The method can be used only by big organizations.

1.9 Framing a Questionnaire or a Schedule

Great care and caution has to be exercised in drafting a questionnaire or a schedule, as this is the basis of collecting information in an investigation. Apart from the care and caution, a lot of expertise and experience of the phenomenon under investigation are required in its preparation. Though there are no hard and fast rules for drafting a questionnaire, still following points should be given due consideration. These points are:

1. **Covering Letter:** In this letter the investigator should introduce himself and make the objectives of the survey clear to the informant. In this, the informant should be assured that the information provided by him will be kept secret. A self-addressed stamped envelope should also be sent, so that the informant can return the questionnaire.
2. **Number of Questions:** The number of questions in the questionnaire should be as small as possible. This results in time saving and is convenient to both, *i.e.,* to the enumerator and the respondent.
3. **Nature of Questions:** Delicate questions should be put with great care. Often indirect questions should be formed to get the answers to these questions.
4. **Simplicity of Questions:** The questions set should be clear, concise, short answered and unambiguous. These should be related to the phenomenon under investigation.
5. **Arrangement of Questions:** There should be a natural and logical order of the questions in a questionnaire. For example, it is not logical to ask a man about his income before asking him about his occupation.
6. **Information Collected is Usable:** It should be noted that the information collected through questions is usable.
7. **Avoid Mathematical Questions:** As far as possible, questions involving mathematical calculations be avoided. It is always better to use multiple choice questions (having four or five alternatives) or simple alternative questions (Yes or No type).
8. **Attractive Layout of the Questionnaire:** The book of the questionnaire should be attractive and questions be suitably spaced for proper answering.

1.10 Collection of Secondary Data

[B.C.A. (Meerut) 2008, 2004; B.C.A. (Lucknow) 2007, 2006, 2004]

Secondary data are those which are collected by some other agency and are used for further investigation. The sources of secondary data can be classified into two:

1. Published Sources
2. Unpublished Sources.

1. **Published Sources**

   Some of the published sources providing secondary data are:

   (i) **Government Publications:** *A* number of government, semi-government and private organizations collect data related to business, trade, prices, consumption, production, industries, income, health and population etc. These publications are very powerful sources of secondary data. Central Statistical Organisation (C.S.O.), National Sample Survey Organisation (N.S.S.O.), Office of the Registrar General and Census Commissioner of India, Directorate of Economics and Statistics and Labour Bureau, Ministry of Labour are a few government publications.

   (ii) **International Publications:** Various governments in the world and international agencies regularly publish reports on data collected by them on various aspects. For example, U.N.O.'s Statistical Year Book and Demography Year Book etc. can be named in this category.

   (iii) **Semi-official Publications:** Local bodies like District Boards, Municipal Corporations publish periodicals providing information about vital factors like health, births and deaths etc.

   (iv) **Reports of Committees and Commissions:** At times State and Central governments appoint committees and commissions with a specific reference to study a phenomenon. The reports of these committees and commissions provide important secondary data. For example, Kothari Commission report on education reforms, report of National Agricultural Commission, Wanchoo Commission report on Taxation and Pay Commission reports etc.

   (v) **Private Publications:** The following private publications may also be enlisted as the source of secondary data:

   (a) **Journals and Newspapers**: Eastern Economists, Monthly Statistics of Trade, Financial Express, Economic Times are some of the Journals and Newspapers which regularly collect and publish data on various aspects of business, economics, commerce and trade.

(b) **Research Publications:** A number of research organizations University departments and institutes like Indian Statistical Institute (I.S.I.) Calcutta and Delhi, I.C.A.R., N.C.E.R.T., I.C.M.R., etc. also contribute significantly to the availability of secondary data.

(c) **Publications of Business and Financial Institutions:** A number of business and financial institutions like Chamber of Commerce and Trade Association, Institute of Chartered Accountants, Sugar Mill Association, Stock Exchanges, Trade Unions and Co-operative societies etc. also contribute significantly for the availability of secondary data in the related areas.

(d) **Articles:** Market reviews and reports also provide data for analysis.

2. **Unpublished Sources:** In some cases data are collected but these are not put in published form. For example research scholars in the institutes and Universities, trade associations and labour bureaus do collect data but they never put it in published form. Still, the data from these sources may be used when needed.

1.11 Limitations of Secondary Data

[B.C.A. (Meerut) 2009]

1. There is a possibility that proper procedure might not have been followed in their collection.
2. These may not be relevant in the present contest.
3. These may not be free from personal bias and prejudices.
4. These may not have the needed accuracy or reliability.
5. These may not be adequate.

1.12 Precautions in Accepting the Secondary Data

[B.C.A. (Agra) 2006]

1. The authority which had collected the data should be reliable and dependable, generally, a government or a semi government agency is more dependable than a private one.
2. The method of collection, including the sampling technique, should be acceptable to the nature of the enquiry in hand.
3. The primary and the secondary purposes should have some uniformity with respect to the aim and objects of the study.
4. The interval between the periods when a data was originally collected to the time when it is going to be used should not be much.
5. The facts on which the original data was collected should remain relevant to the present problem too.

1.13 Advantages of Secondary Data

1. **Economy:** The time, money and labour is saved in collection of secondary data
2. **Reliable:** These are more reliable because these are primarily collected by able and efficient researchers.
3. **Flexibility:** They can be change according to need.
4. **Accuracy:** A high level of accuracy is found in these data because these are collected by efficient persons.
5. **Useful for quick** and generally for quick and regular informations.

Solved Examples

Example 1: *What are the advantages as well as the drawbacks of the study of business problems.*

[B.C.A. (Rohilkhand) 2008; B.C.A. (Agra) 2008, 2005;
B.C.A. (Meerut) 2007, 2006; B.C.A. (Kanpur) 2004, 2002]

Solution: In the mid of 20th, a top American business executive wrote that ''When the history of modern time is finally written, we shall read it as beginning with the age of stream and then progressing through the age of electricity to that of statistics''.

The use of statistical methods in analysis of sales and of accounting records is an old through. In business today, an increasing amount of time and effort is devoted to fact finding operations which costs less of money as well. This is done scientifically by means of the sampling methods. The sampling methods are of particular importance in market research, quality control of products and in forecasting of demand costs. Sample surveys of consumer apptitude toward products and services are routine methods in market research. Manufactures use statistical quality control to reduce the variations in the quality of the finished product. The use of statistical methods helps the business management in the making proper discussions in complex situations. Production, sales, purchase and finance-all needs statistical methods. A business involves quantitative information will sufficient variation in them. So use of statistical methods is a must there.

Drawbacks: The statistical methods being quantitative total for analysis cannot always be consider as ideal. They do not remain the substitutes to the matured common sense statistical methods are only approximates and thus, are not advised to be used where perfect accuracy is wanted. Additionally statistical conclusions are applicable only in the group of observations and not with individual cases.

Example 2: *"Statistics are numerical statements of facts, but all facts numerically stated are not statistics" comment ?*

Solution: This is, rather, a genuine statement. Statistics does not end can not deal with an event or is isolated case. That the height of a particular individual is 5′, Milton was blind. Nepolean was a great warrior are examples of statements that do not come with in the preview of statistics. The wage earned by an individual at a time, taken as an unrelated single unit, is not a statistical datum while wages earned by him over a period of time, taken as a group of related wages, are the different values of the same characteristics and are obviously related and obtain by some repetitive operation. Observation manufactured product illustrates the former while recording of weights of some college students will involved later type of operations.

Example 3: *Elucidate the following statements.*

Solution:

(i) We know that statistics does not deal with a single object or an single event standing along, put it pertains to the characteristics of an aggregate of objects called population. The yields of some specific plot is say 10,000 kgs. It is not for statistical study unless, supplemented by some similar other informations like yields from other identical plots or of same plot recorded in the previous years etc.

(ii) It is not doubt, the first and an indispensable step of a statistical survey that the data collected be reduced or condensed for its subsequents. But statistical operations can not be taken as finished at this stage.

(iii) We know that statistics deals with quantitative or numerically expressed quantitative characteristics. The amount of these characteristics will naturally vary from one object to another thus a set of varying observation remains the subject matter of statistics. There is no other science to deal with such data that is why statistics is branded as a science for studying variations in the variable values.

Example 4: *The science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data" Elucidate it.*

Solution: **Statistics in mathematics applied observational data.** Statistics deal with problems involving empirical or observed evidences about the population characteristic. These evidences, when expressed numerically, are called statistical operations, namely, its treatment, summarisation, analysis and of drawing inference are undertaken. All these steps from the subject matter of statistics and are mathematical in nature. Thus statistics remains the mathematics applied to the collection data. However, the collection of data in itself to is a statistical job.

The science of statistics is a branch of applied mathematics.

The advanced statistical methods are having their roots in the theory of probability which in turn has its foundation in the theory of measure and integration while they are the subject matter of **mathematics.** That is why statistics is taken as branch of applied mathematics. Also the statistical techniques, which are so frequently used in solving the problems of various science of our daily life, are based on different mathematical models and assumptions and remain mathematical in character.

Thus statistical problems embrances the practical application of mathematics involving its problems of high utility and of almost value in everyday science and society.

Problem Set

1. What are the various method of collecting statistical data ? Which one is most reliable and why ?

[B.C.A. (Meerut) 2008, 2006]

2. Distinguish between primary and secondary statistical data and explain the precautions.

[B.C.A. (Meerut) 2001]

3. Classify the methods generally used in the collection of statistical data and state briefly their respective merits and demerits.

[B.C.A. (Delhi) 2004]

4. What are requisites of a good table ?

[B.C.A. (Agra) 2003]

5. Discuss various methods of collecting data. What do you mean by primary and secondary data ?

[B.C.A. (Meerut) 2007, 2001]

6. What do you mean by a questionnaire and a schedule? State the important points to be kept in mind while drafting a questionnaire.

[B.C.A. (Meerut) 2008, 2007, 2006]

7. Define and differentiate between questionnaire and schedule.

[B.C.A. (Meerut) 2007, 2005]

8. Examine critically the important methods of collection of primary data. Can any method be called the best under all circumstances ?

9. What are the various methods of collecting statistical data. Which of these is most reliable and why.

[B.C.A. (Meerut) 2007]

10. Distinguish between primary and secondary data.

[B.C.A. (I.G.N.O.U.) 2011; B.C.A. (Kanpur) 2008, 2007;
B.C.A. (Meerut) 2007, 2006;]

11. Which of the following statements belong to statistics ? Give reasons.
    (i) Indians are not hard workers.
    (ii) I am taller than you.
    (iii) Our birth rate is still a point of worry.

12. Illustrate with suitable examples the use of statistical methods in commerce, business and industry.

13. Write an essay on statistical fallancies.

14. Comment on the following statements:

  (i) Statistics is the science of averages.

  (ii) Statistics is the science of counting.

  (iii) We compute statistics from statistics by statistics.

15. What are the advantages as well as the drawbacks of the study of business problems through statistics.

16. Describe the importance of statistics in trade and commerce.

[B.C.A. (Avadh) 2007]

17. State the brief three important causes of distrust of statistics.

[B.C.A. (Purvanchal) 2006]

18. 90% of those who drink (wine) due before reaching the age of 75 years. Therefore, drinking is bad for longevity. Do you agree ? Analyse.

[B.C.A. (Bundelkhand) 2008]

19. State the four main functions of a statistician.

[B.C.A. (Garhwal) 2011]

20. Define statistics. Describe its functions and importance.

[B.C.A. (A.M.U.) 2008, 2005; B.C.A. (Agra) 2008, 2003]

21. Define statistics and discuss its scope and limitations.

[B.C.A. (Kashi Vidhyapeeth) 2011, 2009; B.C.A. (Meerut) 2010]

22. Explain the importance of statistics in trade and commerce.

[B.C.A. (Agra) 2006]

23. 'Statistics are like clay of which you can made a god or devil as your please' comment.

24. 'Statistics affects everybody and touches like at many point' comment.

[B.C.A. (Purvanchal) 2009; B.C.A. (Rohilkhand) 2009, 2003]

25. ''Statistics is the science of estimates and probabilities''. Give limitations of statistics comment on this statement.

[B.C.A. (Meerut) 2009]

26. Define statistics and discuss its nature and scope.Describe how it is used in social and physical science.

[B.C.A. (Meerut) 2009; B.C.A. (Avadh) 2009, 2006;
B.C.A. (Agra) 2008, 2005]

❑❑❑

Chapter

2

Classification and Tabulation of Data

2.1 Classification

According to **Conner-"*Classification*** *is the process of arranging things (either actually or nationally) in the groups according to their resemblances and affinities and give expression to the unity of attributes that may subsist amongst a diversity of individuals*". **Thus, classification is grouping of data according to their identity, similarity, or resemblances.** For example, students in a class may be grouped in respect of sex, age, marital status etc.

Similarly, letters in the post office are classified according to their destinations *viz.* Delhi, Jaipur, Agra, Kanpur, etc.

2.1.1 Objects of Classification

[B.C.A. (I.G.N.O.U.) 2010, 2004; B.C.A. (Kashi Vidhyapeeth) 2008]

The following are the main objectives of classifying the data:

(i) It condenses the mass of data in an easily assimilable form.

(ii) It eliminates unnecessary details.

(iii) It facilitates comparison and highlights the significant aspect of data.

(iv) It enables one to get a mental picture of the information and helps in drawing inferences.

(v) It helps in the statistical treatment of the information collected.

(vi) It helps in comparative study.

(vii) It helps in clarifying the points of similarity and dissimilarity.

2.1.2 Types of Classification

[B.C.A. (I.G.N.O.U.) 2010, 2004; B.C.A. (Kashi Vidhyapeeth) 2008]

Statistical data are classified in respect of their characteristics. Broadly, there are four basic types of classification namely:

(i) Chronological or Temporal Classification

(ii) Geographical or Spatial Classification

(iii) Qualitative Classification

(iv) Quantitative Classification

(i) Chronological or Temporal Classification: *In* chronological classification *the collected data are arranged according to the order of time expressed in years, months, weeks etc.* The data are generally classified in ascending order of time. For example, the data related with population, sales of a firm, imports and exports of a country are always subjected to chronological classification. Specifically, the estimates of birth rates in India during (1970-79) are:

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Birth Rate	36.8	36.9	36.6	34.6	34.5	35.2	34.2	33.0	33.3	33.0

(ii) Geographical or Spatial Classification: *In this type of classification the data are classified according to geographical region or place.* For instance, the production of paddy in different states in India, production of wheat in different countries etc. For indicating immediate comparison, the observations are either classified in the alphabetical order of the reference places in the order of size of the observation. See the following example:

(a) When names of the countries are in Alphabetical order:

Country	America	China	Denmark	France	India
Yield of Wheat in (kg/acre)	1925	893	225	439	862

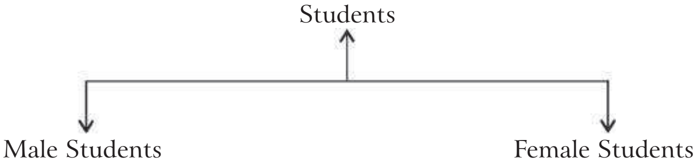
(b) When observations are in descending order:

Country	America	China	India	France	Denmark
Yield of Wheat in (kg/acre)	1925	893	862	439	225

(iii) Qualitative Classification **[B.C.A. (Meerut) 2004, 2002]**

In this type of classification, data are classified on the basis of some attributes or quality like sex, literacy, religion, employment, etc. Such attributes cannot be measured along with a scale. For example, if the students in a class are to be classified in respect of one attribute, say sex, then we can classify them into two classes namely that of males and females.

Similarly, they can also be classified into 'employed' or 'unemployed' on the basis of another attribute 'employment'. Thus, when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. *This type of classification is called* simple or dichotomous classification. A simple classification may be shown as under:



On the other hand, if we classify students simultaneously with respect to two attributes, *e.g.* sex and employment. Then students are first classified with respect to 'sex' into 'Males' and 'Females'. Each of these classes may then be further classified into 'employed' and 'unemployed' on the basis of attribute 'employment' and as such students are classified into four classes, namely:

(i) Male employed, (ii) Male unemployed, (iii) Female employed, (iv) Female unemployed. Still, the classification may be further extended by considering other attributes like marital status etc.

The classification, where two or more attributes are considered and several classes are formed is called **a manifold classification.**

(iv) Quantitative Classification: Grouped Data or Frequency Distribution

In Quantitative Classification, *the collected data are grouped with reference to the characteristics which can be measured and numerically described such as height, weight, sales, imports, age, income, etc.* The first step in the direction of putting observations in some ordered form is to arrange them in ascending or descending order of magnitude. *The data are then said to be in an* array. For example, consider the data consisting of marks secured by 50 students in a class as shown in Table 1:

Table 1: Marks of 50 Students, Ungrouped Data

21	50	42	75	55	67	74	55	47	64
71	61	40	25	25	54	64	37	88	44
31	70	81	51	45	63	49	43	35	67
68	31	38	45	59	75	57	29	66	50
56	84	56	88	63	32	55	88	79	78

Now, the data in above Table 1 may be arranged in an array as shown in Table 2:

Table 2: Data from Table 1 Arranged in an Array

21	31	40	45	51	56	61	66	71	79
25	32	42	47	54	56	63	67	74	81
25	35	43	49	55	57	63	67	75	84
29	37	44	50	55	58	64	68	75	84
31	38	45	50	55	59	64	70	78	88

But still, it does not fully eliminate the confusion which may result from a large number of marks. Thus, to make out data more comprehensive, we group or classify identical observations into ordered class-intervals. *Grouping our data into ordered classes and determining the number of observations in each of the classes, we form a* **Frequency Distribution**. *The tabular form of a frequency distribution gives us a* **Frequency Table.**

2.2 Construction of Frequency Distribution

[B.C.A. (Agra) 2007; B.C.A. (Purvanchal) 2006]

To illustrate the construction of a frequency table, let us again consider the data in Table (1) which represent the marks of 50 students in an examination. Here, as a first step, we decide about the number of classes into which data are to be grouped. This is done arbitrarily, although we are guided by the number of observations to be grouped. Ordinarily, the number of classes should be between 5 and 20. The number of classes would also depend on the number of observations; with larger number of observations one can have more classes. An advantage of a large number of classes is that it leads to better accuracy in the computation of statistical measures like mean etc., but if there are too many classes, it defeats the purpose of grouping. On the other hand, too few classes would be sacrificing too much information and would lead to more errors in calculations for mean etc.

Another factor used for determining the number of classes is the **size, width** or **range** of the class usually called **class-interval** denoted by h. Class-intervals should, to the extent possible, be of uniform size or width. They are easier to comprehend and make the calculations easier. However, there may be exceptions to this rule. The school children, as an example, may be grouped in age classes 6-13, 14-17 and 18-22 of width 8, 4 and 5 respectively. These groups or classes correspond to certain levels of education, *e.g.* elementary, secondary and higher education levels. Further, *the* width of the class should be a whole number conveniently divisible by 2, 3, 5, 10 or 20. Now, after deciding about the number of class intervals, say 10, we calculate the range R (highest marks minus lowest marks) of marks to be grouped. From data in Table 1, the range of marks is

$$R = 88 - 21 = 67$$

Now apply the following formula to get the approximate number of classes in which observations are to be grouped:

$$\text{Number of classes} = \frac{R}{h} = \frac{\text{Range of marks}}{\text{Class} - \text{interval}} = \frac{67}{10} = 6.7 \quad ...(1)$$

Thus, in the grouping of data in Table (1), one should expect to have 6.7 or 7 classes in whole number. Formula (1) can also be used to decide about the class-interval, if we know the range of observations and the number classes, as

$$\text{Class interval} = h = \frac{R}{\text{No. of classes}} \quad ...(2)$$

Having determined the class-interval, one must decide where to start the classes. Since for data in Table 1, the lowest score is 21, so we might begin with 20 as it is common to let the first class start with a number which is a multiple of class interval. Now there are three following ways of describing class limits for different classes.

1. Exclusive Method
2. Inclusive Method
3. True, Actual Class Limits or Class Boundaries

1. **Exclusive Method:** In exclusive method of class formation, we add the class interval 10 to the lower limit of the lowest class to find the upper limit of the class as $20+10=30$. Thus, our lowest class becomes (20–30). The remaining class limits and classes are obtained by adding the class interval to each class limit until we reach the seventh class as (80–90), which contains the highest score of 88 in data in Table 1. The seven exclusive classes thus formed are shown in Table 3:

Table 3: Different Types of Classes

Exclusive Classes	Inclusive Classes	True or Actual Class Limits or Class Boundaries
20—30	20—29	19.5—29.5
30—40	30—39	29.5—39.5
40—50	40—49	39.5—49.5
50—60	50—59	49.5—59.5
60—70	60—69	59.5—69.5
70—80	70—79	69.5—79.5
80—90	80—89	79.5—89.5

Thus, in exclusive method of class formation, *classes are so formed that the upper limit of one class is the lower limit of the next class and therefore, this method of classification ensures continuity between two successive classes which is essential for most of statistical calculations.* However, one should note that in exclusive classes it is presumed that the score or observation equal to the upper limit of the class is exclusive, *e.g.* a score of 30 will be included in the class 30−40 and not in 20−30.

2. **Inclusive Method:** *Unlike exclusive classes,* inclusive classes *include scores or observations which are equal to upper limit of the class.* In the formation of such classes we start with the lower limit 20 of the scores for the first class and then the lowest class is formed as (20–29) so as to include 10 marks (10 being the class interval). These 10 marks are 20, 21, 22, ... 28, 29. The remaining six classes are obtained by adding the class interval to each class limit of the previous class until we get the highest class as (80–89), inclusive classes so obtained are also given in Table 3.

   From the preceding discussion it should be clear that exclusive method should be used when data are of continuous nature or have been measured infractions of unit. Inclusive way of forming classes may be preferred when measurements on the variable are given in whole numbers.

3. **True or Actual Class or Limits or Class Boundaries:** In inclusive method of class formation, it was observed that the upper class limit is not equal to lower class limit of the next class and so there is no continuity between classes. However, for most of statistical calculations it is desirable that classes be continuous. For overcoming this difficulty we assume that an observation does not just represent a point on a continuous scale, but an interval of unit length of which the given observation is the middle point. For example, a score of 20 in an examination represent the interval 19.5 to 20.5 on a continum. Similarly, a score of 29, then, is representable by the interval 28.5 to 29.5. *Thus, the mathematical meaning of an observation is an interval which extends from* 0.5 *unit below to* 0.5 *unit above the face value of an observation on a continum. These class limits of an observation are termed as* true or actual class limits. Thus, the true class limits for the class 20−29 becomes 19.5−29.5. The true limits for other inclusive classes are also shown in Table 3.

   After forming the classes, finally we count the number of observations falling in each class and record the appropriate number in frequently (denoted by f) column. *The number of observations falling in each class is termed as* class frequency. To facilitate counting these frequencies, we also prepare a tally chart with 'tally bar' column as shown in Table 4:

Table 4: Data from Table 1 Group in Frequency Table (Exclusive Method)

Marks Classes	Tally Bar	Frequency (f)
20—30	||||	4
30—40	卌 |	6
40—50	卌 |||	8
50—60	卌 卌 ||	12
60—70	卌 ||||	9
70—80	卌 ||	7
80—90	||||	4
Total		**$N = 50$**

In Table 4, four tally bars are marked for four marks 21, 25, 25 and 29 which lie within the class (20–30), similarly six bars are marked for six scores lying in the class (30–40) and so on. In this way, the raw or ungrouped data in Table 1 are arranged in the form of a frequency distribution showing classes and the corresponding class frequencies. The total of all the frequencies is denoted by N. The frequency table in Table 4 gives the tabular form of this frequency distribution. Similarly, the data in Table 1 can be grouped by using inclusive and true class limits. Such a frequency table is given in Table 5:

Table 5: Raw Data from Table 1 Grouped into Frequency Table (Inclusive Method)

Marks Class	True Class Limits	Class Mid Points	Frequency (f)
20—29	19.5—29.5	24.5	4
30—39	29.5—39.5	34.5	6
40—49	39.5—49.5	44.5	8
50—59	49.5—59.5	54.5	12
60—69	59.5—69.5	64.5	9
70—79	69.5—79.5	74.5	7
80—89	79.5—89.5	84.5	4
Total			$\boldsymbol{N} = \mathbf{50}$

In the process of forming a frequency distribution, we observe that individual observations lose their identity when grouped into classes. We need, therefore, a point to make some assumption about the location of all the observations included in a class. On the assumption that grouped observations are more or less uniformly distributed between class limits, a point mid-way between class limits can be chosen as representative of all the observations in that class. *This point is called the* mid-point *of the class and can be obtained by averaging the lower and upper class limits as follows*:

$$\text{Mid-point of the class} = \frac{\text{lower class limit} + \text{upper class limit}}{2}$$

Thus, the mid-point of the class (20–29) will be (20+29)/2=24.5. The mid-points of the remaining classes are shown in Table 5.

Now, we are in a position to list the steps in grouping a large set of data into a frequency distribution.

2.2.1 Steps in the Construction of Frequency Distribution

(i) *Find the smallest and largest observations in the raw data and find the range R.*

(ii) *Decide on the number of classes.*

(iii) *Divide range by the number of classes to estimate approximate size of the interval.*

(iv) *Find the lower class limit of the lowest class and add to it the class interval to get the upper class limit.*

(v) *Obtain class-limits for the remaining classes by adding the class interval to the limits of the previous class.*

(vi) *Count number of frequencies in each class and check against the total number of observations.*

2.3 Cumulative Frequency Distribution

[B.C.A. (I.G.N.O.U.) 2009; B.C.A. (Kanpur) 2008, 2006]

Sometimes we may be interested to know the number of observations less than or more than a specified value. The principal, for example, may be interested in knowing the number of students scoring above 60 or below 30 marks. Such results can be easily obtained from a cumulative frequency distribution which can be derived from a simple frequency distribution by merging successive class-intervals until all have been cumulated. Such a cumulation procedure may be of two types resulting in two types of cumulative frequencies, namely:

2.3.1 'Less than' Cumulative Frequency Distribution

By adding the frequencies of all observations less than 'true' upper limit of a class, we get 'less than' cumulative frequency. Thus, for calculating these frequencies 'true' upper limit of a class is regarded as the reference point. 'Less than' cumulative frequency table for the grouped data of Table 5 is shown in Table 6:

Table 6: Data from Table 5 Grouped into 'Less than' and 'More than' Cumulative Frequency Table

Marks Classes	'Less than' Cumulative Frequency	Marks Classes	'More than' Cumulative Frequency
Less than 29.5	4 = 4	More than 19.5	50 = 50
Less than 39.5	4+6 = 10	More than 29.5	50–4 = 46
Less than 49.5	4+6+8 = 18	More than 39.5	50–4–6 = 40
Less than 59.5	4+6+8+12 = 30	More than 49.5	50–4–6–8 = 32
Less than 69.5	4+6+8+12+9 = 39	More than 59.5	50–4–6–8–12 = 20
Less than 79.5	4+6+8+12+9 +7= 46	More than 69.5	50–4–6–8–12–9 = 11
Less than 89.5	4+6+8+12+9+7+4=50	More than 79.5	50–4–6–8–12–9–7 = 4

2.3.2 'More than' Cumulative Frequency Distribution

For 'more than' cumulative frequency *'true' lower limit of the class is taken as the reference point.* By adding the frequencies corresponding to more than 'true lower limit' of a class, we get 'more than' cumulative frequency. 'More than' type cumulative frequency table corresponding to data in Table 5 is also given in Table 6.

2.4 Relative Frequency Distribution

Sometimes it is desirable to express the frequencies in a class in the form of proportion or percentage of the total frequency. This may be useful for comparing the distributions of frequencies for two sets of data. The relative frequency *of a class is obtained by dividing the class frequency by the total frequency. A table showing relative frequencies in each class is accordingly called as* relative frequency table. If we multiply each relative frequency by 100, we get a percentage distribution of the frequencies. The relative and percentage frequency table for the data in Table 5 is given in Table 7:

Table 7: Relative and Percentage Frequency Table

Marks Classes	Frequencies (f)	Relative Frequency	Percentage Frequency
20—29	4	.08	$.08 \times 100 = 8$
30—39	6	.12	$.12 \times 100 = 12$
40—49	8	.16	$.16 \times 100 = 16$
50—59	12	.24	$.24 \times 100 = 24$
60—69	9	.18	$.18 \times 100 = 18$
70—79	7	.14	$.14 \times 100 = 14$
80—89	4	.08	$.08 \times 100 = 8$
Total	$N = 50$	**1.00**	**100**

Solved Examples

Example 1: *The marks of 80 students in mathematics are given below. Using inclusive classes, prepare a frequency table by taking the class width as 5.*

84	68	75	82	90	68	88	62	76	93
79	73	88	73	93	60	59	71	85	75
65	61	75	87	62	74	78	95	63	72
78	66	82	75	77	94	74	69	68	60
78	96	89	61	95	75	79	60	83	71
62	79	67	97	85	78	65	76	71	75
80	65	73	57	78	88	76	62	53	74
67	86	73	81	63	72	75	76	85	77

Solution: As a first step in preparing frequency table, we decide the range of marks as R = the highest marks – the lowest marks = 97 – 53 = 44.

Since the lowest mark is 53, we start with the lower limit 50 of the first class so that, with a class width of 5, the lowest class is (50 –54) as it includes 5 marks (50, 51, 52, 53 and 54). The remaining classes can be obtained by adding the class width to each class limit of the previous class until we get the highest class (95 –99) which includes the highest marks 97. Finally, using tally bar, the class frequencies can be recorded as shown in the frequency table.

Frequency Table 8: Showing Distribution of Marks of 80 Students in Mathematics

Classes	Tally Bar	Frequency (f)
50—54	|	01
55—59	||	02
60—64	卌 卌 |	11
65—69	卌 卌	10
70—74	卌 卌 ||	12
75—79	卌 卌 卌 卌 |	21
80—84	卌 |	06
85—89	卌 卌	09
90—94	||||	04
95—99	||||	04
Total		N = **80**

Example 2: *Mid-points in a frequency distribution of age of a group of persons are* 18, 25, 32, 39, 46, 53, 60 *and* 67. *Find*

(i) *the size of the class*

(ii) *the class limits, assuming that the age quoted is the age completed last birthday*

(iii) *the class boundaries* **[B.C.A. (I.G.N.O.U.) 2009, 2006]**

Solution:

Table 9: Showing Class-Limits and Class Boundaries for the given Mid-Point

Mid-Points	Class-Limits	Class-Boundaries
18	15—21	14.5—21.5
25	22—28	21.5—28.5
32	29—35	28.5—35.5
39	36—42	35.5—42.5
46	43—49	42.5—49.5
53	50—56	49.5—56.5
60	57—63	56.5—63.5
67	64—70	63.5—70.5

(i) Since given mid-points have a uniform difference, so class intervals too will be of uniform size. The size of the class is the difference between the mid-values of any two consecutive classes, which is $32 - 25 = 7$ in the present case.

(ii) In case the age quoted is taken as the age completed on last birthday, the variable 'age' assumes only integral values. Thus the 'inclusive classes' of width 7 with above mentioned mid points are to be defined. Therefore, the class-limits for an inclusive class of size 7 having a mid-point 18 will be $(18 - 3, 18 + 3)$ *i.e.* $(15 - 21)$. In the same way, the class-limits for the remaining mid-points can be obtained as shown in the table.

(iii) The class boundaries for different classes can be obtained by adding and subtracting half the size of the class from the corresponding mid-values. Here half of the size of class is 7/2=3.5. Thus, for a class with mid-value 18, the class boundaries will be $(18 - 3.5, 18 + 3.5)$, *i.e.* $(14.5 - 21.5)$. Following the same way, the class boundaries for the remaining classes can be obtained as shown in the table.

Example 3: *Prepare 'Less than' and 'More than' cumulative frequency distributions from the following distribution. Also obtain relative frequency table:*

Wages (₹)	240–250	250–260	260–270	270–280	280–290	290–300
No. of Workers	5	10	18	27	35	5

Solution: 'Less than' and 'More than' cumulative frequency distributions for the above simple frequency distribution are shown in the following table:

'Less than' and 'More than' Cumulative Frequency Distributions

'Less than' Cumulative Frequency Distribution		'More than' Cumulative Frequency Distribution	
Classes	***c.f.***	**Classes**	***c.f.***
Less than 250	5 = 5	More than 240	100 = 100
Less than 260	5+10 =15	More than 250	100 – 5 = 95
Less than 270	5+10+18=33	More than 260	100 – 5 – 10 = 85
Less than 280	5+10+18+27=60	More than 270	100 – 5 – 10 – 18 = 67
Less than 290	5+10+18+27+35=95	More than 280	100 – 5 – 10 – 18 – 27 = 40
Less than 300	5+10+18+27+35+5 =100	More than 290	100 – 5 – 10 – 18 – 27 – 35 = 5

We know that the relative frequency for a class is obtained by dividing the class frequency by the total frequency. The relative frequency distribution so obtained is given in the table below:

The Relative Frequency Distribution of 100 Workers

Wage (₹) Classes	f	Relative Frequency f/N
240—250	5	.05
250—260	10	.10
260—270	18	.18
270—280	27	.27
280—290	35	.35
290—300	5	.05
Total	$N = 100$	**1.00**

Example 4: *In a survey of 50 families, the number of children per family was recorded as under:*

1	3	2	2	0	1	1	2	5	1
2	2	3	3	1	3	2	2	3	2
3	0	0	2	2	5	2	3	2	2
1	5	1	2	3	2	2	3	1	3
4	3	3	5	2	1	0	4	2	1

obtain a discrete frequency table. **[B.C.A. (Avadh) 2006]**

Solution: Here, the variable (X), the number of children, assumes the values 0, 1, 2, 3, 4 and 5. Counting frequencies for each value we get the needed discrete frequency table.

Discrete Frequency Distribution of No. of Children in 50 Families

X No. of Children	f No. of Families
0	4
1	10
2	18
3	12
4	2
5	4
Total	**50**

Example 5: *In a survey, it was found that 50 families bought milk in the following quantities in a particular month. Classify the following data in an individual discrete, continuous and cumulative frequency series:*

19	23	12	21	20
11	7	9	11	14
20	15	22	15	17
10	23	18	11	19
25	22	5	18	17
11	6	21	24	26
8	5	9	20	23
16	22	22	23	13
16	11	16	22	17
23	22	10	7	21

[B.C.A. (Rohilkhand) 2011, 2005. 2003; B.C.A. (Bundelkhand) 2010, 2006, 2004; B.C.A. (Rohtak) 2008, 2004]

Solution:

(i) Individual Series: In the ascending order of items

5	11	16	20	22
5	11	16	20	22
6	11	16	20	23
7	11	17	21	23
7	11	17	21	23
8	12	17	21	23
9	13	18	22	24
9	14	18	22	25
10	15	19	22	25
10	15	19	22	26

(ii) Discrete Series: In the ascending order of the items

Quantity of Milk	Tally Bar	Frequency	Quantity of Milk	Tally Bar	Frequency
5	||	2	16	|||	3
6	|	1	17	|||	3
7	||	2	18	||	2
8	|	1	19	||	2
9	||	2	20	|||	3
10	||	2	21	|||	3
11	𝍸	5	22	𝍸 |	6
12	|	1	23	||||	4
13	|	1	24	|	1
14	|	1	25	||	2
15	||	2	26	|	1

(iii) Frequency Distribution:

Class Interval	Tally Bars	Frequencies
5–10	𝍸 |||	8
10–15	𝍸 𝍸	10
15–20	𝍸 𝍸 ||	12
20–25	𝍸 𝍸 𝍸 ||	17
25–30	|||	3
	Total	$\Sigma f = 50$

(iv) Cumulative Frequency Distribution:

Quantity of Milk	No. of Buyers	Quantity of Milk	No. of Buyers
Less than 5	0	Less than 5	50
Less than 10	8	Less than 10	42
Less than 15	18	Less than 15	32
Less than 20	30	Less than 20	20
Less than 25	47	Less than 25	3
Less than 30	50	Less than 30	0

Example 6: *Prepare a frequency series of the ages of 25 students of class B.B.A. in your school.*

15, 16, 16, 17, 18, 17, 15, 15, 16, 16, 17, 15, 16, 16, 15, 16, 16, 17, 17, 18, 19, 16, 15

Solution: In this question, minimum age is 15 years and the maximum age is 19 years. Therefore, the distribution of frequencies would correspond to the values between 15 and 19.

Frequencies Series

Age	Tally bars	Frequency (f)
15	𝍸 ||	7
16	𝍸 ||||	9
17	𝍸	5
18	|||	3
19	|	1
Total		$\Sigma f = 25$

Example 7: *We have the following data on the monthly expenditure of food (in ₹) for 30 households in a locality.*

115	159	196	205	212	223
256	271	310	129	335	169
184	234	245	241	265	298
144	135	172	173	229	243
220	238	278	243	220	238

(i) Obtain a frequency distribution using following class interval 100–150, 150–200, 200–250, 250–300, 300–350.

(ii) What percentage of households spends less that ₹ 250 per month and what percentage of households spends more than ₹ 200 per month ?

[B.C.A. (Kurukshetra) 2012]

Solution:

(i) The data may be presented in the form of an exclusive series;

Class–interval	**Tally Bars**	**Frequency (f)**
100–150	𝍷𝍷𝍷𝍷	4
150–200	𝍸 𝍷	6
200–250	𝍸 𝍸 𝍷𝍷𝍷	13
250–300	𝍸	5
300–335	𝍷𝍷	2
Total		$\Sigma f = 30$

(ii) No. of households spending less than ₹ 250 per month $= 4 + 6 + 13 = 23$

Percentage of households spending less than ₹ 250 per month $= \frac{23}{10} \times 100 = 76.6\%$

No. of households spending more than ₹ 200 $= 13 + 5 + 2 = 20$

Percentage of households spending more than ₹ 200 $= \frac{20}{30} \times 100 = 66.5\%$

2.5 Three Forms of Data

Going through the process of Quantitative Classification we observe that statistical data can be put in the following three forms:

1. Raw data, ungrouped data or individual series.
2. Discrete frequency distribution.
3. Continuous frequency distribution.

Problem Set-1

1. Define classification and explain various ways of classification.

2. Discuss the importance of classification in statistics. **[B.C.A. (Meerut) 2008, 2007]**

3. Explain the general principles of classification of data for forming a frequency distribution.

4. What is meant by 'Classification' ? State its objectives. Briefly explain different methods of classifying statistical data. **[B.C.A. (Agra) 2006]**

5. Explain the following terms:
   (i) Class-interval
   (ii) Class-frequency
   (iii) Class-limits
   (iv) Class mark or class mid-point
   (v) Frequency distribution
   (vi) Cumulative frequency table
   (vii) Relative frequency table

6. Discuss various steps in the construction of a frequency distribution from raw data. **[B.C.A. (Meerut) 2007, 2005]**

7. Distinguish between:
   (i) Simple and manifold classification
   (ii) Inclusive and exclusive methods of class formation
   (iii) 'Less than' and 'More than' cumulative frequency distributions.

8. Prepare a discrete series from the following data:

62	50	57	58	51	53	62	64	60	61
60	51	64	55	55	52	60	65	58	60
59	52	63	56	56	58	64	63	62	60
58	54	62	54	54	60	65	60	62	59
65	56	63	52	53	62	53	61	61	59

9. If class mid-points in a frequency distribution of a group of persons are 25, 32, 39, 46, 53, 60, 67, 74 and 81, find
   (i) Size of class interval, and (ii) The class boundaries.

10. Classify the following by taking class-intervals such that their mid-values are 17, 22, 27, 32 and so on:

30	42	30	54	40	48	15	17	51	42
25	41	30	27	42	36	28	26	37	54
44	31	36	40	36	22	30	31	19	48
16	42	32	21	22	46	31	41	21	16

11. What are the objects of classification of data ? Discuss the different methods of classification.

12. Distinguish between classification and tabulation. Discuss the purpose, methods and importance of classification.

13. Find the frequency table having grades of wages with class interval of Rs. 20 each from the following data of daily wages received by 30 workers in a certain factory daily wages in rupees are :

140, 160, 140, 220, 130, 150, 240, 120, 230, 140, 200, 170, 210, 220, 180, 180, 190, 200, 270, 160, 150, 110, 120, 210, 200, 170, 180, 190, 230

Answers-1

8.

x	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	Total
f	1	2	3	3	3	2	3	1	4	3	7	3	6	3	3	3	60

9. (i) 7

(ii) 21.5–28.5, 28.5–35.5, 35.5–42.5, 42.5–49.5, 49.5–56.5, 56.5–63.5, 63.5–70.5, 70.5–77.5 77.5–84.5

10.

Class	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	Total
Frequency	5	4	4	8	4	9	3	3	40

2.6 Tabular Presentation

2.6.1 Table and Tabulation

[B.C.A. (Meerut) 2009]

Tabulation is the *process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information.* A table is *a systematic arrangement of classified data in columns and rows.* Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious. Classification and Tabulation as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation, data are classified and then displayed under different columns and rows of a table.

2.6.2 Advantages of Tabulation

Statistical data arranged in a tabular form serve following objectives:

(i) It simplifies complex data and the data presented are easily understood.

(ii) It facilitates comparison of related facts.

(iii) It facilitates computation of various statistical measures like averages, dispersion correlation etc.

(iv) It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided. Moreover, the needed information can be easily located.

(v) Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

2.6.3 Preparing a Table

[B.C.A. (Meerut) 2009]

The making of a compact table is itself an art. This should contain all the information needed within the smallest possible space. What the purpose of tabulation is and how the tabulated information is to be used, are the main points to be kept in mind while preparing for a statistical table. An ideal table should consist of the following main parts :

(i) **Table Number:** A table should be numbered for easy reference and identification. This number, if possible, should be written in the centre at the top of the table. Sometimes it is also written just before the title of the table.

(ii) **Title:** A good table should have a clearly worded, brief but unambiguous title explaining the nature of data contained in the table. It should also state arrangement of data and the period covered. The title should be placed centrally on the top of a table just below the table number (or just after table number in the same line).

(iii) **Captions or Column Headings:** Captions in a table stand for brief and self-explanatory headings of vertical columns. Captions may involve headings and sub-headings as well. The unit of data contained should also be given for each column. Usually, a relatively less important and shorter classification should be tabulated in the columns.

(iv) **Stubs or Row Designations:** Stubs stands for brief and self-explanatory headings of horizontal rows. Normally, a relatively more important classification is given in rows. Also a variable with a large number of classes is usually represented in rows. For example, rows may stand for score classes and columns for data related to sex of students. In the process, there will be many rows for score classes but only two columns for male and female students.

(v) **Body:** The body of the table contains the numerical information of frequency of observations in the different cells. This arrangement of data is according to the description of captions and stubs.

(vi) **Footnotes:** Footnotes are given at the foot of the table for explanation of any fact or information included in the table which needs some explanation. Thus, they are meant for explaining or providing further details about the data, that have not been covered in title, captions and stubs.

(vii) **Sources of Data:** Lastly one should also mention the source of information from which data are taken. This may preferably include the name of the author, volume, page and the year of publication. This should also state whether the data contained in the table is of primary[1] or secondary[2] nature.

A model structure of a table is given below:

Model Structure of a Table

← Table Number →

← Title of The Table →

Stub Heading	**Caption Headings**	**Total**
	Caption Sub-Headings	
← Stub Sub-Heading →	Body	
Total		

Foot Note : 1. 2.

Source Note : 1. 2.

2.6.4 Requirements of a Good Table

A good statistical table is not merely a careless grouping of columns and rows but should be such that it summarizes the total information in an easily accessible form in minimum possible space. Thus, while preparing a table, one must have a clear idea of the information to be presented, the facts to be compared and the point to be stressed. Though, there is no hard and fast rule for forming a table yet a few general points should be kept in mind.

(i) A table should be formed in keeping with the objects of a statistical enquiry.

(ii) A table should be carefully prepared so that it is easily understandable.

(iii) A table should be formed so as to suit the size of the paper. But such an adjustment should not be at the cost of legibility.

(iv) If the figures in the table are large, they should be suitably rounded or approximated. The method of approximation and units of measurements too should be specified.

(v) Rows and columns in a table should be numbered and certain figures to be stressed may be put in 'box' or 'circle' or in bold letters.

(vi) The arrangement of rows and columns should be in a logical and systematic order. This arrangement may be an alphabetical; chronological or according to size.

(vii) The rows and columns are separated by single, double or thick lines to represent various classes and sub-classes used. The corresponding proportions or percentages should be given in adjoining rows and columns to enable comparison. A vertical expansion of the table is generally more convenient than the horizontal one.

(viii) The averages or totals of different rows should be given at the right of the table and that of columns at the bottom of the table. Totals for every sub-class too should be mentioned.

(ix) In case it is not possible to accommodate all the information in a single table, it is better to have two or more related tables.

2.6.5 Type of Tables

Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:

(i) **Simple or One-Way Table:** A simple or one-way table *is the simplest table which contains data of one characteristic only.* A simple table is easy to construct and simple to follow. For example, the adjacent blank table may be used to show the number of adults in different occupations in a locality.

The Number of Adults in Different Occupations in a Locality

Occupations	No. of Adults
Total	

(ii) Two-Way Table: *A table which contains data on two characteristics is called a* **two-way table**. In such case, therefore, either stub or caption is divided into two co-ordinate parts. In the given table, as an example, the caption may be further divided in respect of sex. This sub-division is shown in the adjacent two-way table which now contains two characteristics namely, occupation and sex.

The Number of Adults in a Locality in respect of Occupation and Sex

Occupation	No. of Adults		Total
	Male	Female	
Total			

(iii) Manifold Table: Thus, more and more complex tables can be formed by including other characteristics. For example, we may further classify the caption sub-headings in the above table in respect of 'marital status', 'religion' and 'socio-economic status' etc. *A table in which more than two characteristics of data are considered, is called a* manifold table. For instance, table below shows three characteristics, namely, occupation, sex and marital status.

The Number of Adults in a Locality in respect of Occupation, Sex and Marital Status

Occupations	No. of Adults						Total
	Male			Female			
	M	U	Total	M	U	Total	
Total							

Foot note — 'M' stands for married and 'U' for unmarried.

Manifold tables/though complex, are good in practice as these enable full information to be incorporated and facilitate an analysis of all related facts. Still, as a normal practice, not more than four characteristics should be represented in one table to avoid confusion. Other related tables may be formed to show the remaining characteristics.

Example 8: *Prepare a blank table to show the distribution of students in a college according to marks, sex and two sections.*

Solution: Here three characteristics are to be simultaneously represented in a table and so either a row or a column heading may be sub-divided to show the third characteristics. As a rule, sub-division of a column heading is advisable. The needed blank table thus takes the form.

Distribution of Students according to Marks, Sex and Section

Marks Classes	Boys			Girls			Total
	Sec. A	Sec. B	Total	Sec. A	Sec. B	Total	
0—10 10—20 20—30 30—40							
Total							

Example 9: *In a sample study about coffee drinking habits in two towns the following information was recorded:*

Town A *'Females were 40%, total coffee drinkers were 45% and male non-coffee drinkers were 20%.* ***Town B:*** *Male were 55%, male non-coffee drinkers were 30% and female coffee drinkers were 15%. Represent the data in a tabular form.*

Solution: In such questions, we first form a blank table showing the needed classification. Then we fill up the cells with given frequencies which have been shown by bold letters in the following table. The remaining frequencies can be easily obtained by adjusting the marginal totals. For example, in town A, the male percentage can be obtained by subtracting female percentage from 100, *i.e.* $100-40=60$. Similarly, the percentage of male coffee drinkers can be obtained by subtracting the percentage of male non-coffee drinkers from 60, *i.e.* $60-20=40$. The remaining frequencies in other cells are also obtained in the same.

Way-Table Showing the Coffee Drinking Habits of *Town A* and *B*
(Figures in Percentage)

Sex / Drinking Habit	*Town A*			*Town B*		
	Male	Female	Total	Male	Female	Total
Coffee drinkers	40	05	45	25	15	40
Non-Coffee drinkers	20	35	55	30	30	60
Total	60	40	100	55	45	100

Example 10: *Prepare a blank table showing the distribution of students of a college according to age, sex, class and residence for arranging physical training.*

Solution: Table showing the distribution of students of a college according to age, sex, class and residence:

		Students					
Class	Age group	Male			Female		
		Rural	Urban	Total	Rural	Urban	Total
B.B.A	15–20						
	20–25						
	25–30						
B.C.A	15–20						
	20–25						
	25–30						
M.B.A	15–20						
	20–25						
	25–30						
Total							

2.7 Difference between Classification and Tabulation

[B.C.A. (A.M.U.) 2011]

The main objective of both classification and tabulation is to present the collected data in a summarised and compact form but there are few differences which are as follows:

1. **Difference in Sequence:** From the view of sequence in statistical inquiry, classification is done first and tabulation later *i.e.,* classification forms the basis for tabulation.
2. **Difference in Presentation:** In classification, data are classified into different classes according to their similarities and dissimilarities, while in tabulation the classified data are placed in columns and rows. Thus, tabulation is an mechanical function of classification.
3. **Difference in Method:** Classification is a method of statistical analysis, while tabulation is a method of presentation of data.
4. **Basis of Presentation:** In classification, data are classified into classes and sub-classes while in tabulation data are presented under headings and sub-headings.
5. **Use of Data:** In classification only original data are used, while in tabulation derivatives like percentages.

Problem Set-2

1. Define table and tabulation. **[B.C.A. (Meerut) 2008, 2006]**
2. Discuss the importance of tabulation. **[B.C.A. (Rohilkhand) 2009, 2008, 2006]**
3. State clearly the essentials of a good table.
4. Write a note on various parts of an ideal table. **[B.C.A. (Agra) 2007, 2005]**
5. Define 'Caption/and Stubs'.
6. Define a two-way table with example.
7. Define a manifold table with example.
8. Distinguish between 'Classification and Tabulation'. **[B.C.A. (Meerut) 2008, 2007, 2006]**
9. Prepare a blank table to show the number of employees in a firm according to:
   - (i) Sex: male and female
   - (ii) Age-Groups: below 25, 25–35, 35–45 and above 45.
10. Construct a blank table showing the distribution of workers in a factory according to:
    - (i) Religion: Hindus, Muslims, others.
    - (ii) Age Groups: 0–30, 30–50, 50 and above.
    - (iii) Skill: Skilled and Unskilled.
    - (iv) Sex: Male and Female.
11. Write a note on importance of classification and tabulation. **[B.C.A. (Meerut) 2008]**
12. What is statistical table ? Mention of rule of construction of a table. **[B.C.A. (Avadh) 2009, 2007]**

❑❑❑

Chapter 3

Frequency Distribution and Graphical Representation

3.1 Display Data

There are numerous ways of displaying statistical data pictorially. *Some of the main forms of such visual presentation are*:

1. Diagrammatic Presentation
2. Graphical Presentation
3. Pictures
4. Maps

3.1.1 Diagrammatic Presentation

In a **diagrammatical presentation** of statistical data, we make use of geometric figures like bars, squares, rectangles, circles etc. Data classified according to some chronological order, geographical region or attributes is generally presented by **diagrams and pictures**. **Maps** are generally used for geographical presentation of statistical data. **Graphical presentation,** *a subject matter of the next chapter, is used for the pictorial representation of a frequency distribution or a cumulative frequency distribution.*

3.1.1.1 Importance of Diagrams and Graphs

(i) They are attractive and impressive.

(ii) They are even appealing to a common man.

(iii) They facilitate comparison.

(iv) They have universal acceptability.

(v) They are useful for determining some statistical measures of data like median, mode, quartiles and other partition values.

3.1.1.2 Rules for Constructing Diagrams

There are following rules for constructing diagrams:

(i) Every diagram should be suitably and briefly titled.

(ii) It should suit to the size of the paper.

(iii) It should be neat and attractive.

(iv) It should be clearly indexed.

(v) Diagrams should contain foot notes and a proper spacing between the figures.

(vi) The details in the diagram should be self explanatory.

3.1.1.3 Type of Diagrams

(A) One-Dimensional Diagrams

One-dimensional diagrams also called **Bar-diagrams**, are most widely used diagrams for the visual presentation. Bar-diagrams of are the following types:

(i) Simple Bar-Diagrams
(ii) Multiple Bar-Diagrams
(iii) Sub-Divided Bar-Diagrams
(iv) Percentage Bar-Diagrams
(v) Deviation Bar-Diagrams
(vi) Broken Bars

(i) Simple Bar-Diagrams: A bar is a thick whose thickness is taken only for making the diagram attractive, Bar-diagrams are called **one-dimensional** as only the length of the bar represent the magnitude of the value of the variable. Thus, a bar diagram consists of a number of rectangles and is used only for one-dimensional comparisons.

Solved Examples

Example 1: *Draw the bar-diagram to represent the following data related to a school:*

Year	1990	1991	1992	1993	1994	1995
No. of Students	210	242	290	315	340	355

Present the data with a suitable diagram. **[B.C.A. (Meerut) 2011]**

Solution: In this case data can be represented by a bar-diagram as the figures relate to one variable only for bar-diagram see Fig. 3.1.

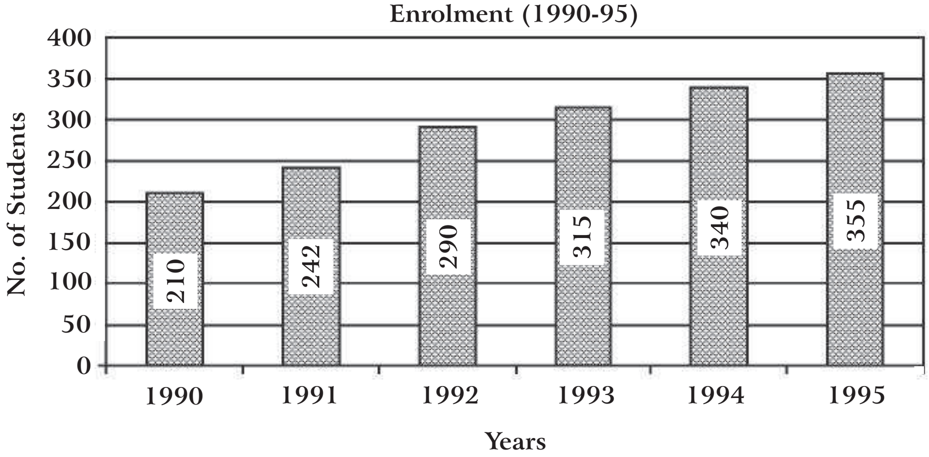


Fig. 3.1: Single Base Diagram

✦ **Limitations of Bar-Diagram**

A bar-diagram has the following limitations:

(a) This diagram should be used only when the given figures do-not vary much and the proportion between the figures is small.

(b) This diagram can be used to represent data on one variable only.

(ii) Multiple Bar-Diagrams: A Multiple-bar diagram *is used when a comparison is to be made between two or more variables. These are also used for comparing magnitudes of one variable in two or three aspects.* The technique of drawing such diagrams is the same as that of simple bar-diagrams. The difference is that in the present case a group of bars representing is placed side by side. The bars are distinguished by shading or colours to show the representation of different variable. The following is an example of multiple-bar diagram.

Example 2: *Following data relate to the faculty-wise enrolment of students in a college:*

Years	1993	1994	1995
No. of Arts Students	95	110	120
No. of Science Students	160	170	165
No. of Commerce Students	75	95	110

Represent the data by a suitable diagram.

Solution: In this example, year-wise data on three aspects (Arts, Science and Commerce) over years is given. So, a multiple bar-diagram should be used for the purpose. See Fig 3.2.

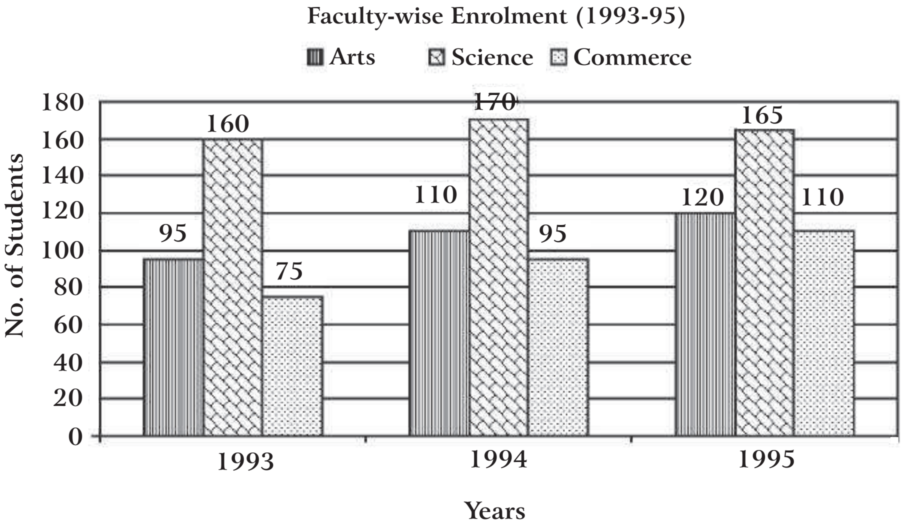


Fig. 3.2: Multiple-Bar Diagram

(iii) Sub-Divided Bar-Diagram: *The* sub-divided bar-diagram, *also known as* components bar-diagram, *is useful in a situation when it is necessary to show and compare the breakup of one variable into several components.* In this case, bars are drawn to represent the total or aggregate magnitudes of the variable; one bar to represent each time period or geographical area. Each bar is then sub-divided in several parts, each part representing a component of the aggregate or total. Different shades or colours are used to distinguish between the components. Various components of the variable are shown in the same order in different bars for easy comparison; the largest component is usually indicated on the lower portion of the bar and the smallest component on its uppermost portion. For example, the total number of students in a college in different years, classified according to sex, may be represented by a sub-divided bar-diagram. Following example will further clarify the purpose and the procedure of drawing a sub-divided bar-diagram.

Example 3: *Following data relate to year-wise enrolment in a college, classified according to sex:*

Year	**1990–91**	**1991–92**	**1992–93**	**1993–94**	**1994–95**
No. of Girls	810	825	844	780	820
No. of Boys	1215	1160	1325	1410	1480
Total	**2025**	**1985**	**2169**	**2190**	**2300**

Represent the data by a suitable diagram.

Solution: The sub-divided bar-diagram to represent the given data is shown in Fig. 3.3.

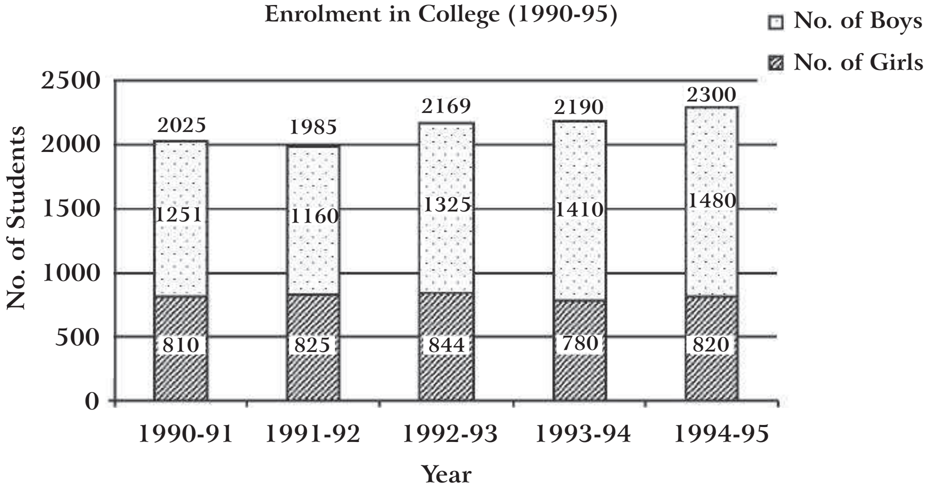


Fig. 3.3: Sub-divided Bar Diagram

(iv) **Percentage Bar-Diagram:** *The construction of percentage bar-diagram is similar to the sub-divided bar chart.* The difference between the two is that, in the sub-divided bar diagram, the component parts are shown in absolute quantities, while in the percentage bar diagram, the component parts are transformed into percentages of the total. The advantage of this diagram is that all the bars are drawn with equal heights representing a total of 100. These bars are then sub-divided in terms of percentages of the components.

Example 4: *Represent the data in Example 2 by a percentage bar-diagram.*

[B.C.A. (I.G.N.O.U.) 2006]

Solution: For representing the data by a percentage bar-diagram, we obtain the percentage of students faculty-wise, for each year as shown in the Table 1.

Table 1: Percentage of Students, Faculty-wise and for Each Year

Year	Percentage of Students			Total %
	Commerce	Science	Arts	
1993	22.73	48.48	28.79	100.00
1994	25.33	45.34	29.33	100.00
1995	27.85	41.77	30.38	100.00

The data in Table 1 is represented by a percentage bar-diagram in Fig. 3.4.

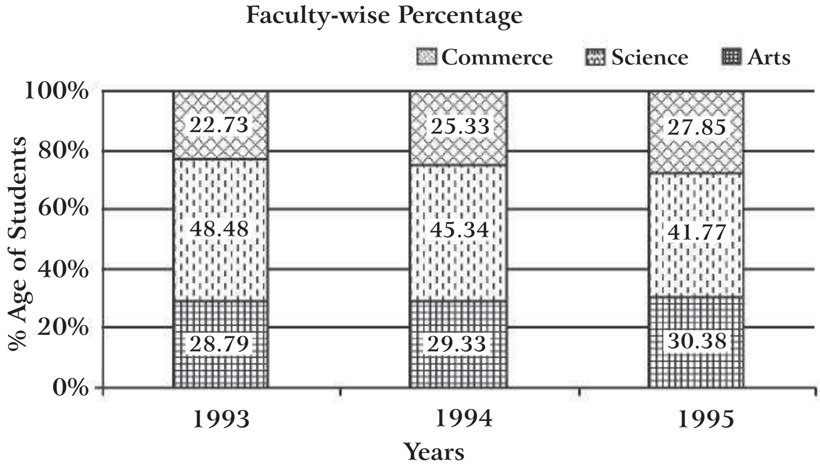


Fig. 3.4: Percentage Bar-Diagram

(v) Deviation Bar-Diagrams: *Deviation bar-diagrams are used to show the net magnitudes of a phenomenon, i.e. net profit, net loss, net exports or imports, etc.* Bars in these diagrams can assume both negative and positive values. Positive valued bars are shown above the base line (x-axis) while negative valued bars are drawn below it. The concept has been clarified with the following example.

Example 5: *Depict the following data by a suitable diagram:*

(Balance of Trade = Export - Import)

Year	*Export*	*Import*	*Balance of Trade (Millions ₹)*
1994	110	140	– 30
1995	115	96	+19
1996	120	100	+20

[B.C.A. (I.G.N.O.U.) 2006]

Solution: Deviation bar-diagram representing data on trade balance is shown in Fig. 3.5.

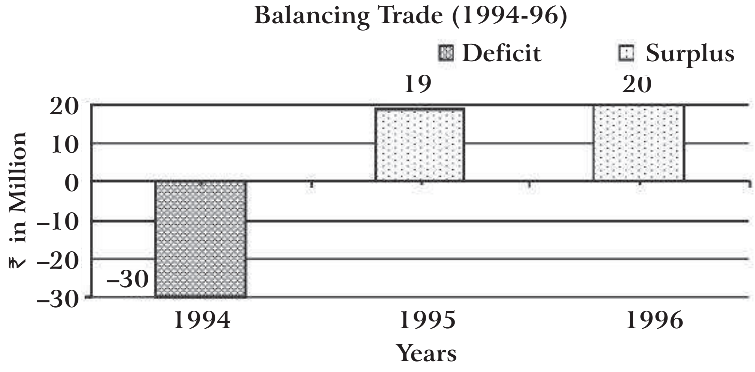


Fig. 3.5: Deviation Bar-Diagram

(vi) **Broken Bars:** Sometimes we wish to represent series having wide variations in values, *i.e.* some values may be very small and others very large. *In such a situation difficulty arises in choosing a suitable scale to accommodate smaller and larger values together. The difficulty is overcome by using* **broken bars** *for representing the larger values.*

Example 6: *The following data relate to sales in five firms A, B, C, D and E:*

Firms	*A*	*B*	*C*	*D*	*E*
Sales (In lakhs ₹)	25	38	300	200	56

Use a suitable bar-diagram to represent the data.

Solution: Since there is a big gap between the smaller and the larger scale values, so broken bars may be used to represent the larger values as shown in Fig. 3.6:

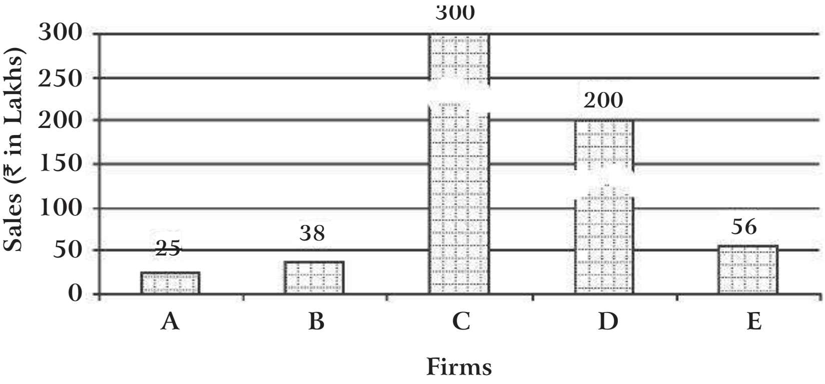


Fig. 3.6: Broken-Bar Diagram

(B) **Two-Dimensional Diagrams or Area Diagrams:** In one-dimensional diagrams we take only the length of the bars into consideration and width is taken for the beautification of the diagrams. *However, in* two-dimensional diagrams *or* area diagrams, *the length as well as the breadth of the geometric figure is taken into consideration.* Such diagrams are also useful in situations when the proportion between the magnitudes of given values of the variable is quite large. Some of the popular two-dimensional diagrams are:

(i) Rectangle Diagrams

(ii) Squares and Circle Diagrams

(iii) Pie-diagrams

(i) **Rectangle Diagrams:** These diagrams are used for two-dimensional comparisons. While drawing the rectangles, we consider the length as well as the width of the bars. *Thus, rectangles used for area diagram vary in heights as well as in the width, so that areas of the rectangles represent the magnitudes of the variable over time or space, or over some other characteristic of variation.*

Rectangles are generally used for representing the total expenditure of two or more families made on various items. These are also used for making comparisons of expenses on various items of production and profit earned per-unit in a number of firms. The following example will clarify.

Example 7: *The following data represent the expenditure of the two families on various items:*

Represents the data by a rectangle diagram.

S.No.	*Items*	*Expenditure* (₹)	
		Family A	*Family B*
1.	*Food*	1,200	1,700
2.	*Clothing*	500	800
3.	*House Rent*	600	900
4.	*Fuel and Electricity*	250	300
5.	*Miscellaneous*	450	800
	Total	**3,000**	**4,500**

Solution: In this example, the width of the rectangles is kept in proportion to the total expenditure of the families, *i.e.* 3000 : 4500 or 2 : 3 and the height show the sub-divisions of the expenditures. See the Fig. 3.7.

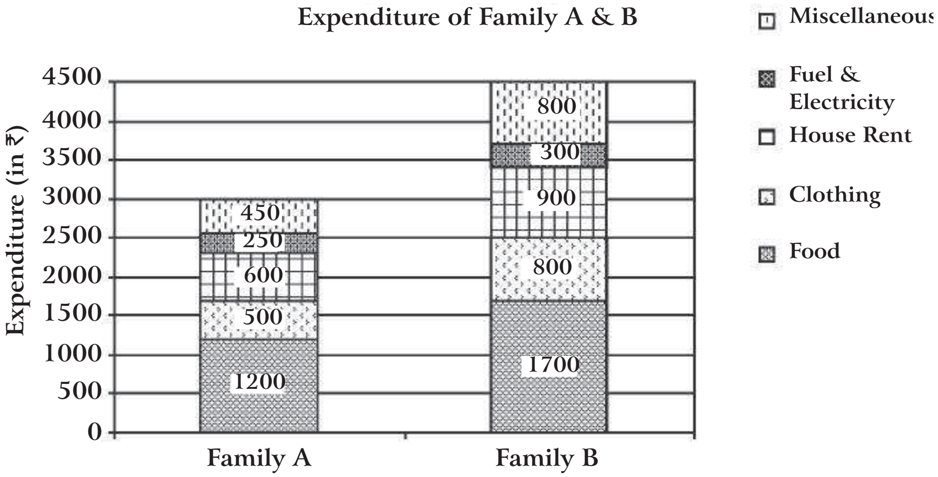


Fig. 3.7: Rectangle Diagram

Example 8: *Represent the data in example 7 by a rectangle diagram on percentage basis.*

Solution: For representing the data by a rectangle diagram on percentage basis, we express the expenditure.

On various items as percentage of the total expenditure for both the families as shown in Table 2:

Table 2: Family Expenditures (in Percentages)

S.No.	Items	Family A		Family B	
		₹	%	₹	%
1.	Food	1200	40.00	1700	37.78
2.	Clothing	500	16.67	800	17.77
3.	House Rent	600	20.00	900	20.00
4.	Fuel and Electricity	250	8.33	300	6.68
5.	Miscellaneous	450	15.00	800	17.77
	Total	**3000**	**100.00**	**4500**	**100.00**

Again the widths of the rectangles are kept in proportion to the total expenditure, *i.e.* 2 : 3 and expenditure is now shown in percentage and not in absolute terms. See the Fig. 3.8.

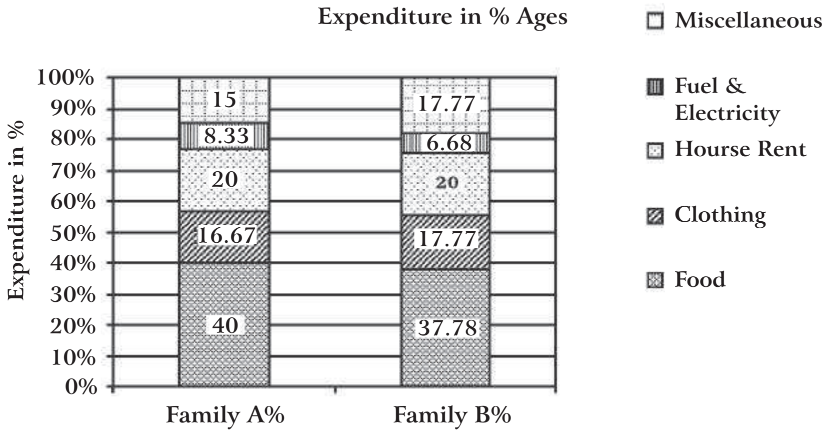


Fig. 3.8: Rectangular Diagram

(ii) Squares and Circle Diagrams: Rectangle diagram is not very useful when the proportion between the magnitudes of the given values is quite large. For example, if the total expenditure of the two families in example 7 are 45,000 and 3000, then the width of the rectangles will be in the ratio 15 : 1 which is an inconvenient ratio to show diagrammatically. To overcome this situation, we use **squares** or **circle** diagrams.

The procedure for drawing square or a circle diagram is quite simple. For drawing squares, sides of the squares are kept proportional to the magnitudes of the values and for circle diagrams, radii of the circles should be proportional. An example for such diagram will clarify.

Example 9: *The following data relate to the plan outlay of a country for three plans:*

Three Years Plan	*I*	*IV*	*VII*
Outlay (₹ '000 Crores)	196	2060	8820

Represent the data by square diagram.

Solution: As there exist a very large proportion between the outlays of different plans, square diagram is suitable. For drawing a square diagram, we first calculate the square roots of the given values as shown in **Table 3:**

Table 3

Plan	Outlay (₹ '000 Crores)	Square Root	Side of the Square (in cms)
I	196	14.00	0.70
IV	2060	45.39	2.26
VII	8820	93.91	4.70

Here the sides of the squares are in the ratio of the square roots of the outlay. The square diagram is shown in Fig. 3.9 next page.

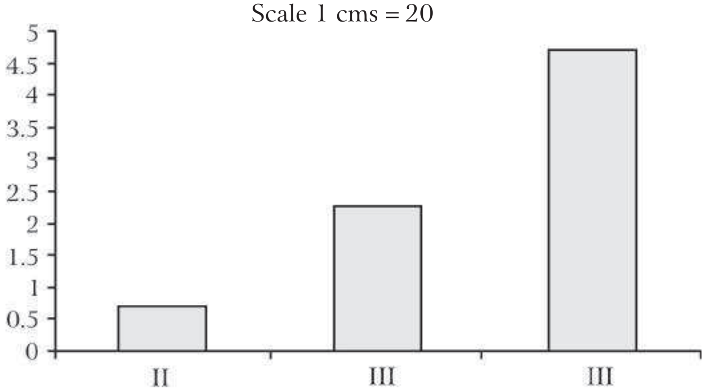


Fig. 3.9: Plans Square Diagram

Example 10: *Represent the data in example 9 by a circle.*

[B.C.A. (Kurukshetra) 2011, 2005]

Solution: In this case the radii of the circles will be kept proportional to the magnitude. The circle diagram is shown in Fig. 3.10:

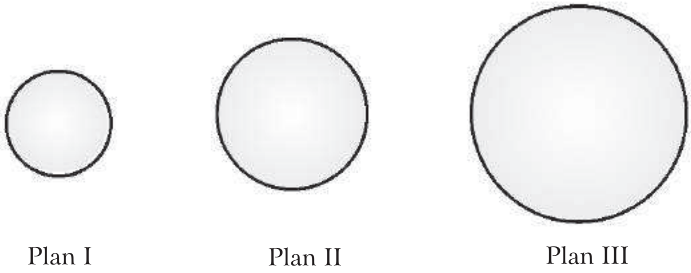


Fig. 3.10: Plans Circle Diagram

(iii) Pie-Diagrams: Pie-diagrams are most popular diagrams used for representing breakdown of an aggregate into its components or sub-divisions. Thus, this diagram is generally used to compare the relations between various sub-divisions of a value. Pie-diagram is *a circle divided into sectors with areas equal to the corresponding components. A* pie-diagram shows the components or sub-divisions in terms of percentages only and not in absolute terms. The areas of the sectors are proportional to the angles subtended at the centre of *the* circle. For example, a pie-diagram may be used for representing India's tax revenue from various sources like excise and other taxes, income tax, corporate taxes etc. Steps in the construction of a Pie-diagram.

(a) Plot a circle with its radius in proportion to the aggregate of the given values.

(b) Convert the component values in terms of degrees in the circle taking the aggregate of the values as 360°.

(c) For simplicity, it is desirable to arrange the component values in some logical pattern. For example, these may be arranged according to their sizes.

(d) Mark the points on the circle representing the size of each sector using a protractor.

(e) Represent each sector with different identification schemes.

The procedure will be clear from the following example:

Example 11: *The following data relate to faculty-wise enrolment in a college:*

Faculty	*Science*	*Arts*	*Commerce*	*Total*
No. of Students	2,010	1,100	2,390	5,500

Represents the data by a pie-diagram.

Solution: First we convert the given values into corresponding degrees in the circle taking the total number of students 5,500 as equal to 360°. The computation is shown in **Table 4:**

Table 4: Calculation for a Pie-Diagram

Faculty	No. of Students	Angles (in Degrees)
Science	2010	$\frac{2010}{5500} \times 360 = 131.56$
Arts	1100	$\frac{1100}{5500} \times 360 = 72.00$
Commerce	2390	$\frac{2390}{5500} \times 360 = 156.44$
Total	**5500**	**360.00**

Pie-diagram representing data is shown in Fig. 3.11.

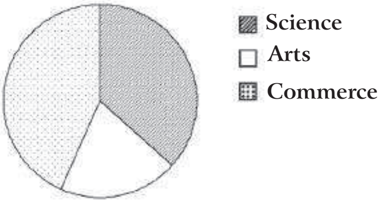


Fig. 3.11: Pie-Diagram

✦ **Important Remark**

A multiple pie-diagram is used for two-dimensional comparisons where a variable value is shown over time, space or in terms of some other characteristics and the variable values are also broken into components. The following example will clarify.

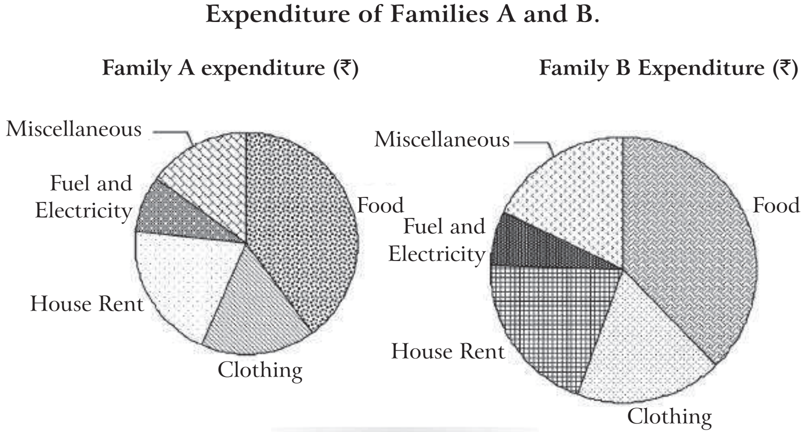
Example 12: *Represent the data in example 7 by a multiple pie-diagram.*

Solution: In this case also we convert the expenditure on various items for both the families into corresponding degrees in the circle as shown in **Table 5:**

Table 5: Computation for a Multiple Pie-Diagram

Items of Expenditure	Family A		Family B	
	Exp. (₹)	Angle (in Degrees)	Exp. (₹)	Angle (in Degrees)
1. Food	1200	$\frac{1200}{3000}\times 360 = 144.0$	1700	$\frac{1700}{4500}\times 360 = 136.0$
2. Clothing	500	$\frac{500}{3000}\times 360 = 60.0$	800	$\frac{800}{4500}\times 360 = 64.0$
3. House Rent	600	$\frac{600}{3000}\times 360 = 72.0$	900	$\frac{900}{4500}\times 360 = 72.0$
4. Fuel and Electricity	250	$\frac{250}{3000}\times 360 = 30.0$	300	$\frac{300}{4500}\times 360 = 24.0$
5. Miscellaneous	450	$\frac{450}{3000}\times 360 = 54.0$	800	$\frac{800}{4500}\times 360 = 64.0$
Total	3000	360.0	4500	360.0

A multiple pie-diagram is shown in Fig. 3.12. Here the radii of the two circles are in the ratio 3 : 2.


Expenditure of Families A and B.

Family A expenditure (₹) **Family B Expenditure (₹)**

Fig. 3.12: Pie-Diagram

(C) Pictorial Diagrams or Pictograms: Statistical data may be represented with the help of pictures also. Such a presentation is called pictorial diagram or pictogram. In pictograms, the magnitude of the values are explained with the help of pictures. The method is widely used by the government and private organisations. The method is specially useful when the trends in data are to be conveyed to the general public not having much of the mathematical background. In a pictogram, a symbolic picture represent a fixed quantity or magnitude of the variable. Obviously, the number of symbols represents the total magnitude of the values.

(D) Cartograms: Statistical data classified according to geographical regions are also representable with the help of suitable maps. *The representation of statistical data by maps is called cartogram.* For example, the data showing region-wise measurements of rainfall, region-wise density of population, etc. can be shown on a map of the region. Different type of shades or colours may be used for representing the magnitudes of the variable.

✦ **Limitations of Diagrammatic Representation**

Diagrams are used to have a quick knowledge of the pattern of data. However, these cannot be considered as a substitute for classification or tabulation. When compared with the tabular form of data, diagrams have the following limitations:

(i) Data on a large number of variables cannot be presented in a diagram while a tabular form may include the same.

(ii) Diagrams presents a vague comparison between various magnitudes of a phenomenon and it is generally not exact.

(iii) Diagrams are useful for conveying the message or pattern of data to a layman while an expert can draw the necessary conclusions from the tabular form of data itself.

3.1.2 Graphical Presentation

There are many forms of graphs which can be broadly classified as:

1. Graphs of the Frequency Distribution
2. Graphs of the Time Series or Line Graphs

3.1.2.1 Graphs of the Frequency Distribution

The graphs representing a frequency distribution are:

(i) Histogram

(ii) Frequency Polygon

(iii) Frequency Curve

(iv) Cumulative Frequency Curve or ‘Ogive’.

✦ The System of Co-ordinate Axes; General Principles

Graphing is done with reference to two perpendicular lines called co-ordinate axes. The vertical line (*yy′*) is called *y* axis and the horizontal (*xx′*) as *x*-axis. The point of their intersection is called the origin. Fig. 3.13 represents a system of co-ordinate axes.

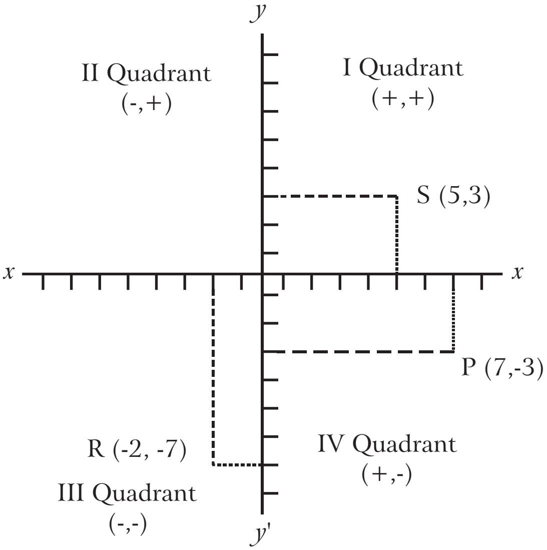


Fig. 3.13: The System of Coordinates Axes

In Fig. 3.13 the origin, denoted by O, is the zero point or reference point for the two axes. All distances measured along the *x*-axis to the right of O are called positive and those to the left of O negative. Similarly, distances measured along the *y*-axis above the origin are positive, and below it negative. Four divisions resulting by the intersection of two axes are called **quadrants.** In the upper right division or first quadrant, both *x* and *y* measures are positive (+, +). In the upper left division or second quadrant, *x* is negative and *y* positive (–, +). In the lower left division or third quadrant, both *x* and *y* measures are negative (–, –); while in the lower right or fourth quadrant, *x* is positive and *y* negative (+, –). These four quadrants have been marked in figure *The distance of a point from origin along x axis is called* ***abscissa****; and the distance of the point from origin on the y axis the* **ordinate**.

(i) **Histogram:** In such graphs the class-intervals are plotted along *x*-axis and frequencies along *y* axis. The first class-interval is taken on *x*-axis and on it a rectangle with the frequency corresponding to that class interval as height is erected. Then the second class-interval is taken on *x*-axis and on it a rectangle with the frequency corresponding to this second class-interval as height is erected. In this way rectangles corresponding to each of the given class-intervals are erected side by side.

The figure so formed is known as **histogram**. Here the area of rectangles are directly proportional to their corresponding class frequencies.

Note 1: In order to draw a histogram, it is necessary that the class-intervals are given by exclusive method. If these are given by inclusive method, then these should be first of all converted into exclusive method.

Note 2: If the given class intervals are not equal, the frequencies are divided by the lengths of the corresponding class-intervals the heights of the columns or rectangles are taken proportional to these and ratios so that the areas of these rectangles are proportional to the respective frequencies.

Example 13: *Construct a histogram to represent the following data:*

Class-Interval	0—10	10—20	20—30	30—40	40—50	50—60
Frequency	25	35	70	55	30	20

Solution:

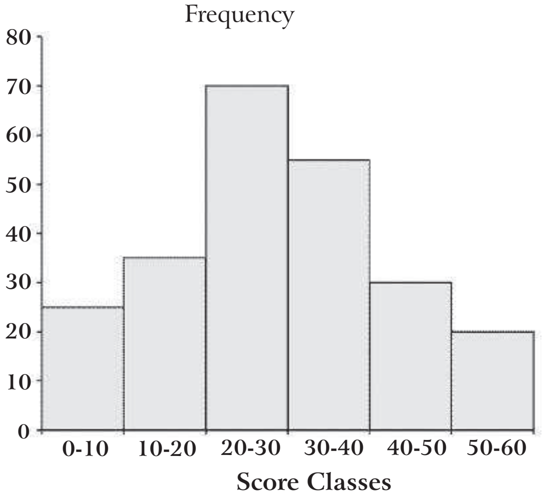


Fig. 3.14: Histrogram

Take the class intervals on x-axis and frequencies on y-axis. Since the frequency for the class-interval 0–10 is 25, so erect on the line depicting 0–10 on x axis as base a rectangle of height 25. Similarly erect on the line depicting the class interval 10–20 on x axis as base a rectangle of height 35 and so on. Thus we erect six rectangles corresponding to given six class intervals. The Fig. 3.14, so formed the required histogram.

✦ Histogram (Unequal Class-Interval)

To take an example, let us arbitrarily group classes 30–39 and 40–49 into one class as 30–49 in Table 6.

Table 6: Frequency Distribution

Scores Classes	True Class Limits	Mid Points	Frequency (*f*)
20—29	19.5—29.5	24.5	4
30—39	29.5—39.5	34.5	6
40—49	39.5—49.5	44.5	8
50—59	49.5—59.5	54.5	12
60—69	59.5—69.5	64.5	9
70—79	69.5—79.5	74.5	7
80—89	79.5—89.5	84.5	4
Total			$N = 50$

Similarly, classes 60–69 and 70–79 are grouped as 60–79. Grouped classes and corresponding frequencies thus obtained are shown in Table 7.

Table 6, Data in Table 7, Grouped into Unequal Class-Intervals

Scores Classes	True Class Limits	Class Interval	Class Frequency	Height of Rectangles
20—29	19.5—29.5	10	4	4/1=4
30—49	29.5—49.5	20	14	14/2=7
50—59	49.5—59.5	10	12	12/1=12
60—79	59.5—79.5	20	16	16/2=8
80—89	79.5—89.5	10	4	4/1=4
Total			$N = 50$	

In Table 7 the class-interval of second and fourth class is twice that of the rest of the classes. Thus, the frequencies in second and fourth classes are not comparable with other classes. To establish this comparability, the frequencies in the larger classes should be halved or divided by two. Thus, before forming histogram for frequency distribution with unequal class-intervals, all larger classes should be expressed as multiples of smaller classes; and then divide the corresponding class frequencies by these multiples. This division then, gives the height of rectangles (as given in the last column of Table 7) to be formed over given classes. Now, if the class-interval 10 of the class 19.5 – 29.5 is taken as one unit then class interval 20 of the class 29.5–49.5 will be equal to 2 units. Thus, a rectangle of height 7 formed over 29.5–49.5 will represent an area $7 \times 2 = 14$ which is, in accordance with the principle of histogram, equal to class frequency. Similarly, the height of the rectangle to be formed over class 59.5–79.5 will be 8 so that its area $8 \times 2 = 16$ will represent the class frequency. However, heights of other rectangles formed over classes of unit lengths will remain equal to corresponding class frequencies. The histogram obtained for the frequency distribution of scores in Table 7 is given in Fig. 3.15.

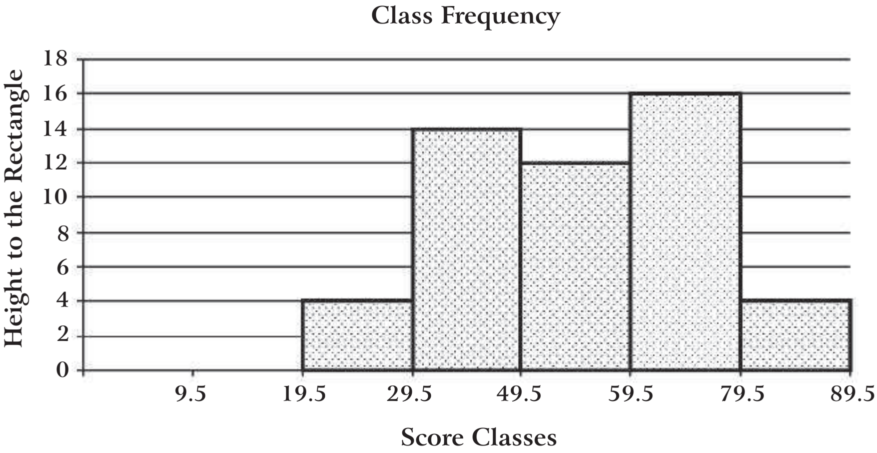


Fig. 3.15: Histogram

We have seen that histogram is a graphical representation of a frequency distribution. It is supposed to represent all the characteristics of a good table. Histogram too should be given a clear, brief and self-explanatory title. Index showing the scales of measurements on both axes should be given in the upper right corner of the figure and units of measurements should be mentioned along the axes so as to make the graph easily readable.

(ii) **Frequency Polygon:** In these polygons, the value of the variate is assume to be concentrated at the mid-value of the interval. To construct this polygon, the mid-value of the class-intervals are taken as abscissae or x-co-ordinates and the corresponding frequencies are ordinates or y co-ordinates and these points are plotted on paper. These points are then joined in order by straight lines. Finally the first of these points is joined to the lower limit of the first interval and the last of these points to the upper limit of the last interval. The polygon so formed by these lines and x-axis is called **frequency polygon.** The defect of frequency polygon is due to fact that the value of the variate is assumed to be concentrated at the mid-value of the class-interval which is generally not true.

Example 14: *Construct the frequency polygon to represent the following data:*

Class-Interval	0—10	10—20	20—30	30—40	40—50	50—60
Frequency	25	35	70	55	30	20

[B.C.A. (Meerut) 2006]

Solution: Find out the mid-values of the intervals *i.e.* 5, 15, 25, 35, 45, 55 and take them as x co-ordinates and the frequencies 25, 35, 70, 55, 30, 20 of the class interval $(0-10), (10-20) \ldots (50-60)$ as y-coordinates and thus we have the points (5, 25), (15, 35), (25,70), (35, 55), (45, 30), (55, 20). Plot these points and join them in order by straight lines. Then join first of these points (5, 25) to (0, 0) as 0 is the lower limit of the first class-interval 0-10 and join last of these points (55, 20) to (60, 0) as 60 is the upper limit of last class interval 50–60.

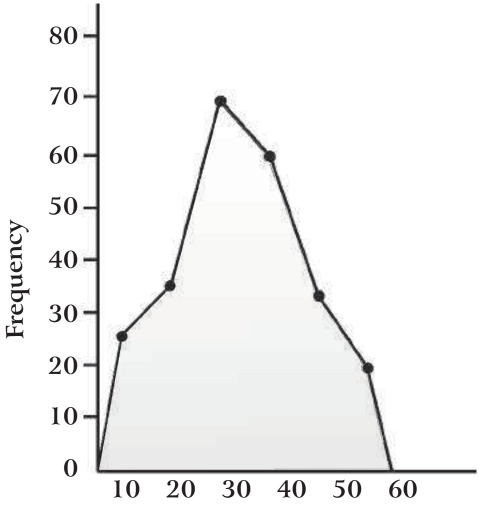


Fig. 3.16: Frequency Polygon

The polygon so obtained is the required frequency polygon. See the Fig. 3.16.

Aliter: Here we mark the mid-points of the tops of each rectangle or column of the histogram and join them in order by straight lines. Then join first of the above marks with the point depicting the lower limit of the first interval and the last of the above marks with the point depicting the upper limit of the interval. The polygon so obtained in the required frequency polygon.

(iii) Frequency Curve: If in a group distribution, the width of the class-intervals become very small then the number of observations increases and the frequency polygon tends to a smooth continuous curve which is called **frequency curve.** To draw the frequency curve, we plot the points as in the case of frequency polygon taking the mid-points of the class-intervals as abscissae and corresponding frequencies as ordinates. Then a free hand curves is drawn through these points which is known as frequency curve. See the Fig. 3.17 next page.

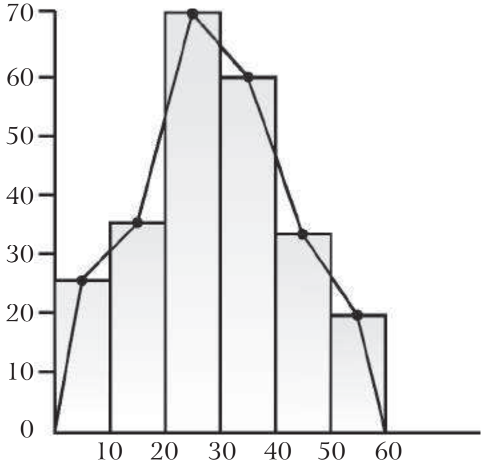


Fig. 3.17: Frequency Polygon

The frequency curve corresponding to the data given example 14 is given in the adjoining Fig. 3.18.

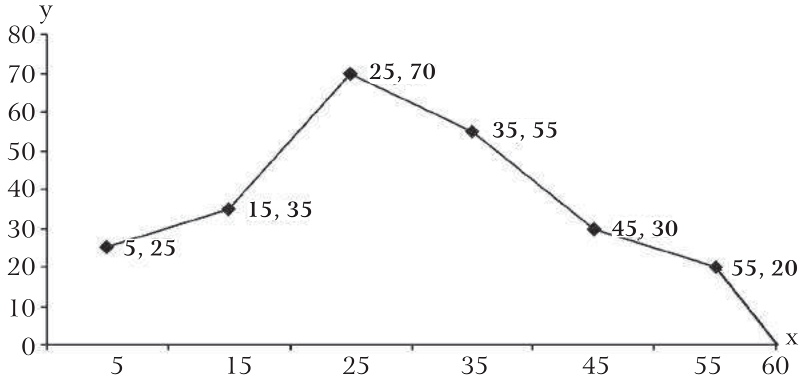


Fig. 3.18: Frequency Curve

(iv) Cumulative Frequency Curve or 'Ogive'

[B.C.A. (Kanpur) 2008, 2005; B.C.A. (Meerut) 2004, 2002]

A cumulative frequency curve or 'Ogive' is a graphical representation of a cumulative frequency distribution. The 'less than' or 'more than' cumulative frequency tables in Table 8 gives either the number.

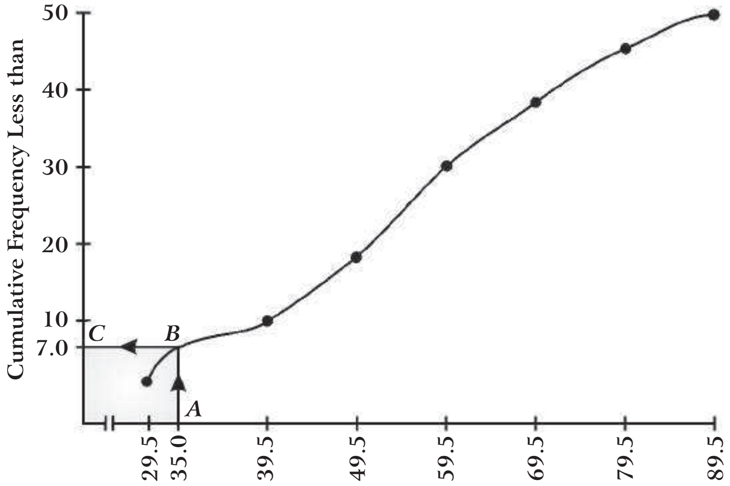
Table 8 'Less than' and 'More than' Cumulative Frequency Table for the Data in Table 6.

Score Classes	Less than C.F.	Score Classes	More than C.F.
Less than 29.5	4	More than 19.5	50
Less than 39.5	10	More than 29.5	46
Less than 49.5	18	More than 39.5	40
Less than 59.5	30	More than 49.5	32
Less than 69.5	39	More than 59.5	20
Less than 79.5	46	More than 69.5	11
Less than 89.5	50	More than 79.5	4

of students scoring below the upper class limits or above the lower class limits. But these tables fail to provide the number of students scoring below or above the intermediate score within a class. For example, the cumulative frequency table in Table 8 shows that 10 students have scored below 39.5 and 46 have scored above 29.5, but it fails to give the number of students scoring below or above 35 which is an intermediate score in class 29.5–39.5. We may further need, for example, the number of students in various grades, say A, B or C where grade A is given to a score of 65 or more, B for above 45 and below 65 and C for the rest. Though, this sort of information could be obtained by using arithmetic interpolation between class limits, yet it can be obtained from ogives with less effort and sufficient accuracy. Ogives are also drawn for determining median, a measure of central tendency, as will be discussed in the First chapter. There are two forms of ogive depending on the types of cumulative frequencies, namely—

(a) **Less than 'Ogive':** In the construction of 'ogives' also, we use true class limits. In the formation of less than 'ogive', upper true limits of the classes are plotted along the horizontal axis and the corresponding less than cumulative frequencies are marked along the vertical axis. The points so obtained are then joined by a free hand smooth curve to get less than 'ogive' as given in Fig. 3.19, for the cumulative frequency distribution in Table.

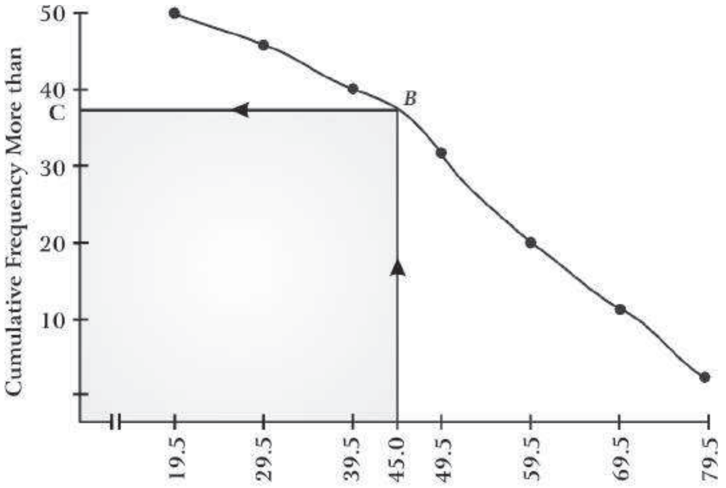
The ogive in Fig. 3.19 can now be used to get number of students scoring below a given score. For example suppose the number of students scoring below 35 is to be estimated. Then 35 is first marked on horizontal axis as a point say A and a vertical line drawn through this point cuts the ogive at some point, say B. This cutting point is then projected on vertical axis to a point C. The point C, when measured on vertical axis, gives the required number of students as 7.



Under True Class Limits Ogive Less than, Cumulative Frequency

Fig. 3.19

(b) **More than 'Ogive':** For more than 'ogive' true lower limits of the classes are plotted along the horizontal axis and the corresponding more than cumulative frequencies are marked along the vertical axis. The points are then joined by a smooth free hand curve to give more than 'ogive' as shown in Fig. 3.20. The cumulative frequency distribution in Table 8 has been used for the purpose.



Ogive More than; Cumulative Frequency

Fig. 3.20

More than 'ogive' in Fig. 3.20, can now be readily used to find the number or percentage of students scoring above a given score. For example, in estimating the number of students scoring above 45, we first mark 45 on the horizontal axis as A.

Now, a vertical line through this point cuts the ogive at point B. This point B is then finally projected to vertical axis as point C. This point C, when measured on vertical axis, gives the required estimated number of students as 36.

Example 15: *Discuss the role of diagrams in presenting statistical data.*

Or

Explain the utility of diagrammatic presentation of data.

[B.C.A. (M.D.U.) Rohtak 2010, 2009, 2005, 2003;
B.C.A. (Lucknow) 2010; B.C.A. (Rohilkhand) 2008, 2007]

Solution: In order to provide an attractive form to the classified data, these are presented as tables, diagrams and graphs. Although the complex data are presented in an easy form through rows and columns under tabulation, yet these fail to attract the people because a special technical knowledge is required to understand them. Moreover, it is the best medium of comparison between two or more items, Data can be presented in diagrammatic form only if these are also available in tabular form.

Example 16: *Explain the rules, types and limitations of presentation of data.*

[B.C.A. (Lucknow) 2005, 2007; B.C.A. (Agra) 2002, 2004]

Solution: The following rules must be followed while making a diagrammatic presentation of data, so that, it may made attractive:

1. **Suitable Title:** A suitable title is necessary to make the diagram attractive. The title must be brief and clear, so that, the objective of diagram may become clear Just by having a view of it.
2. **Size of Diagram:** The size of diagram must also be taken care of while presenting the data in the form of a diagram. The size of diagram should be in accordance with that of the paper.
3. **Use of Colours and Sings:** Various colours and sings must be used in order to provide an attractive form to the diagram. A symbol chart should be drawn on the right side of the diagram.
4. **Scale:** An appropriate scale should be determined before presenting data in the form of diagrams. The scale must be neither very big nor very short.
5. **Limited use of Words and Numericals:** Least possible use of words and numerical must be made while drawing diagrams.
6. **Types of Diagrams:** Diagrams are of various types like diagram, pie diagram, rectangular diagram etc. it is not necessary that every diagram can be used in each and every situation. So, it is necessary to select the appropriate diagram well in advance.
7. **Presentation of Importants Facts:** Use of diagrams is an attractive method of presentation of important facts. With their help, the important facts can be presented in an attractive form by making use of symbols, colours etc. so that these facts may be easily conveyed to the public.

8. **Use of Graph Paper:** Graph papers can also be used while presenting data in the form of diagrams. The diagrams can be given more attractive and simpler form through the graphs.

9. **Clarity:** While presenting diagrams. Special attention should be paid to their clarity. Much information must be include in a single diagram as it will be difficult to understand such a diagram.

   Thus diagrams can be made more attractive by paying attention to the above given facts.

3.1.2.2 Types of Diagrams

Many diagrams are used to present data. Their explanation is as follows:

1. **One Dimensional Diagram:** These are also called **'Bar Diagrams'**. These diagrams are widely used. In these diagrams, are widely used. In these diagrams, the height is given more importance than the breadth or thickness of the diagram. Under it, the data are presented in the form of bars or rectangles. Any person can understand these diagrams easily. One dimensional or bar diagrams may be of the following types:

   (i) **Line Diagram:** When the data are also presented through simple lines only, then these are called line diagrams.

   (ii) **Simple Bar Diagram:** These are also prepared like line diagrams. The difference between the two is that while preparing it, much attention is provided to breadth in order to make the diagram more attractive.

   These are used to demonstrate only a single item.

   (iii) **Multiple Bar Diagram:** When two or more variables are depicted in a single diagram, then it is known as multiple bar diagram. These diagram have a great role to play in Comparative study.

   (iv) **Sub-Divided Bar Diagram:** Sub-divided bar diagrams are used in a case when the sub-parts of two or more variables are also given. Under it, all bar diagrams are divided in different parts and various colours and signs are used to show the difference.

   (v) **Percentage Bar Diagram:** In this method, first of all, the values of variables are depicted in length of each variable is assumed to below. Every part is presented in different colours or symbols.

   (vi) **Deviation Bar Diagram:** These are also called **'Dio-dimensional Bar Diagrams'**. These diagrams are used to show the deviations with reference to time and situation. Deviation is divided into two parts; (i) Positive Deviation and (ii) Negative Deviation. Positive deviation are shown above x-axis and Negative deviation are shown below x-axis.

2. **Two Dimensional Diagrams:** The length and breadth are paid special attention in these diagram. The bars represent the data. These diagrams have the following forms:

   (i) **Rectangular Diagram:** Both the length and breadth are paid special attention in these diagrams. These are also known as 'Area Diagrams'.

   (ii) **Percentage Rectangular Diagram:** Under it firstly, the items are depicted in percentage form and then, in the form of rectangle. The length of bar is assumed to be 100. There after, the rectangular diagram is prepared and sub-divided.

   (iii) **Square Diagram:** The length and breadth of such diagrams are equal. These diagrams are used mainly in such situations when the number of items is more and it is not possible to present them in the form of bar diagrams.

   (iv) **Circle or Pie Diagram:** Under it, a circle is divided into different part and the items are presented in these parts in percentage form.

3. **Three-Dimensional Diagrams:** These diagrams are used when the number of item is very large and it is not possible to present them even through Pie or Square diagram. In these diagrams, each of length, breadth and thickness has special importance.

Thus, it is clear that data can be presented through diagrams in many ways. An appropriate diagram should be selected on the basis of nature of the data.

✦ Limitations of Diagrammatic Presentation

The description of limitations of diagrammatic presentation is as follows:

(i) **Misuse of Diagrams:** It is possible to misuse the information presented diagrammatically because if the diagrams are not prepared properly, then they will not present the right information.

(ii) **Difficulty in Analysis:** Data are not presented in diagrammatical form when it is essential to make their comparative study but the diagrammatically presented data can not be analysed. It is very difficult to make a final conclusion on their basis.

(iii) **Not Useful for Researcher:** The diagrammatic presentation of data is not useful for the researchers because it does not provide any help to the researchers in the analysis of data.

(iv) **Helpful in Estimation Only:** The complete information can not be obtained by presenting the data diagrammatically. It is helpful only in providing an estimation of data. On the contrary, complete information can be easily obtained if the data is presented in a tabular form.

(v) **Not Helpful in Presenting Much Information:** It is not possible to make use of diagrams when a lot of information is to be presented. Only the simple and small data can be presented by using diagrams.

(vi) **Presentation of Approximate Values:** The numerical data can not be presented through diagrams, as a result of which, these lack accuracy and it also becomes difficult to obtain useful information. Only an approximate value of data can be obtained through diagrams.

(vii) **Complexity of Diagrams:** Some diagrams are very complex like two-dimensional and three dimensional diagrams. Due to complexity, these diagrams can not be easily understood by public. A special skill is required to understand them.

(viii) **More Difference in Data:** If there is a lot of difference in data, then it is almost impossible to present them through diagrams. For example, if the data are in a ratio of 6 : 8000. Then their scale con not be prepared.

Thus, it is clear from the above that various limitations of diagrams are also there along with their uses.

Example 17: *What is a graphic presentation of Data? Discuss its merits and demerits.*

[B.C.A. (A.M.U.) 2009]

Solution: Graphic presentation is also an attractive method of presenting the data. Although the data related to time is presented though diagrams. Yet in order to present frequency distribution, time series etc. mainly graphs are used. Graph papers are used for it. The items depicted on graph paper are simple, clear, accurate and more attractive.

✦ Merits of Utility of Graphic Presentation

The explanation of main merits of graphic presentation is as follows:

(i) **Simple and Attractive:** Data can be presented in a simple and attractive manner through graphs, as a result of which, these can be easily understood by public.

(ii) **Helpful in Estimation:** With their help, even the future data can be easily estimated, For example with the help of data about sales from 1982 to 1992. The sales in the year 2000 can be easily estimated.

(iii) **Helpful in Comparison:** Graphs are also helpful in making comparison in the items related to various time.

(vi) **More Memorable:** Due to large number of data, it is very difficult to remember them, but it becomes easy if the data are presented graphically because graphs have a longer image on people mind.

(v) **Estimation of Correlation:** With their help, the correlation between two data can be estimated by presenting them on a single graph.

(vi) **Knowledge of Frequency Distribution:** The knowledge of the nature of frequency distribution can be easily obtain with their help. Frequency distribution can be depicted in various forms like histogram, frequency polygon, simple frequency curve, cumulative frequency curve etc.

✦ Demerits of Graphic Presentation

(i) **Lack of Proof:** The graphically presented data can not be used as proof because these lack accuracy. Their only objective is to proses the complex data in a simpler form, so that, people may understand them easily.

(ii) **Difficult to Understand:** The data presented through graph is difficult to understand as only the person having technical knowledge can understand then so, it is difficult for ordinary man to understand these graphs.

(iii) **Difficult in Analysis:** The graphically presented data can not be used as proof because these lack accuracy. Their only objective is to present the complex data in a simple form, so that people may understand them easily.

(vi) **Wrong Conclusions:** The conclusions drawn from graphs are not accurate and generally, wrong indications are provided by them because a little change in criterion may lead to a great change in line.

(v) **Provides Inadequate Information:** The graphic presentation of data provides inadequate information only.

(vi) **Misuse of Graphs:** There may be various misused of graphic presentation of data because if the graphs are not prepared properly, then these will not provide right information.

Example 18: *Represent the following data by mean of Histogram:*

Weekly Wages	10–15	15–20	20–25	25–30	30–40	40–60	60–80
No. of Workers	7	19	27	15	12	12	8

[B.C.A. (Avadh) 2010, 2008]

Solution: Since the class intervals are unequal frequency are to be adjusted.

Computations of New Frequency of Frequency Density

Week of Wages	No. of Workers	Adjustment	New Frequency
10–15	7	5/5=1	7
15–20	19	5/5=1	19
20–25	27	5/5=1	27
25–30	15	5/5=1	15
30–40	12	10/5=2	12 ÷ 2=6
40–60	12	20/5=4	12 ÷ 4=3
60–80	8	20/5=4	8 ÷ 4=2

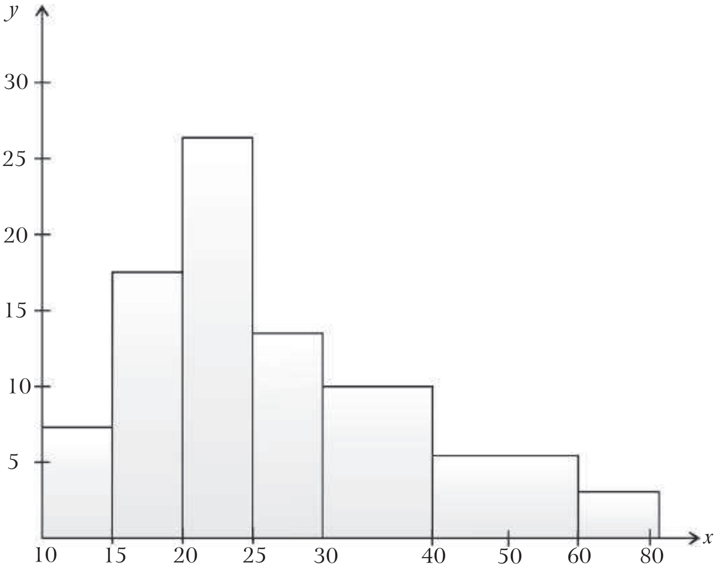


Fig. 3.21: Histogram

Example 19: *Present the following data diagrammatically:*

	←Firm A→	←Firm B→	←Firm C→
Raw Material (₹)	4	5	4
Labour (₹)	3	3	4
Misc Expenses (₹)	1	2	3
Total Cost (₹)	8	10	11
Sales Price (₹)	10	10	10
Profit	+2	–	–1

[B.C.A. (Purvanchal) 2008; B.C.A. (A.M.U) 2005]

Solution:

←Selling prize on % age basis→

Particulars	Firm A		Firm B		Firm C	
	Actual	%	Actual	%	Actual	%
Raw material (₹)	4	40	5	50	4	40
Labour (₹)	3	30	3	30	4	40
Misc expenses (₹)	1	10	2	20	3	30
Total cost (₹)	8	80	10	100	11	110
Sales price (₹)	10	100	10	100	10	100
Profit or Loss	+2	+20%	–		–1	–10%

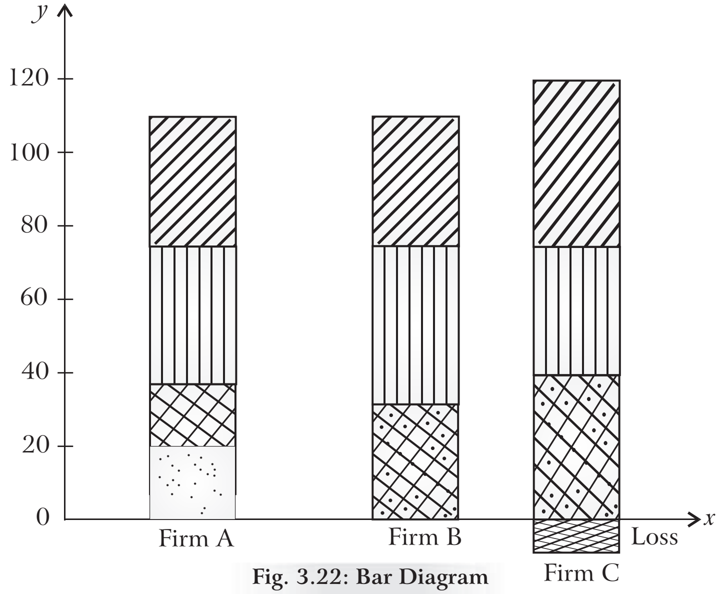


Fig. 3.22: Bar Diagram

3.2 Graphs of the Time Series

A series of observations recorded over time is called a **time series**. For example data related to the population of a country recorded at ten-yearly intervals, annual production of crops, the wholesale price index over a number of months, the daily closing price of a share on the stock exchange. *In fact, the data related with business and economic activities in general, recorded over time gives rise to a* **time series**. Obviously the variations in a time series are most important to an economist. The variations in the time series can also be presented graphically. These graphs are called **line graphs.**

Steps in the construction of line graphs.

1. x-axis is used as time scale, *i.e.* year, months or days are marked along x-axis.
2. Variations in the variable are to be marked along y-axis on a suitable scale.
3. The marked points are joined by straight lines to show the pattern of variation. The following example will clarify:

Example 20: *Represent the following data by a suitable line graph. (Rate per* 1000 *population per annum)*

Year	1911–21	1921–31	1931–41	1941–51	1951–61	1961–71	1971–81
Birth Rate	48.1	46.4	45.2	39.9	41.7	41.1	37.1
Death Rate	47.2	36.3	31.2	27.4	22.8	18.9	14.8
Natural Growth Rate	0.9	10.1	14.0	12.5	18.9	22.2	22.3

Solution: The line graph is shown in Fig 3.23:

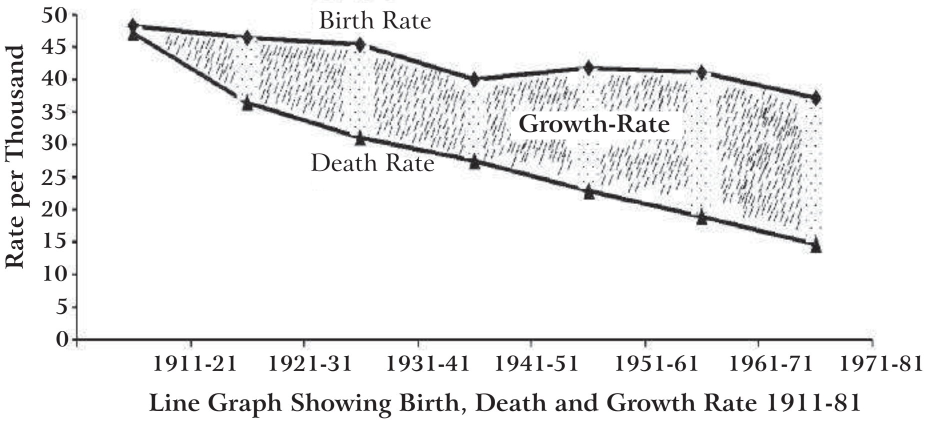


Fig. 3.23: Line Graph

Example 21: *Represent the given data by a histogram.*

Class	30–35	35–40	40–45	45–50	50–55	55–60	60–65
Frequency	1	5	22	14	11	6	1

[B.C.A. (Bundelkhand) 2006]

Solution:

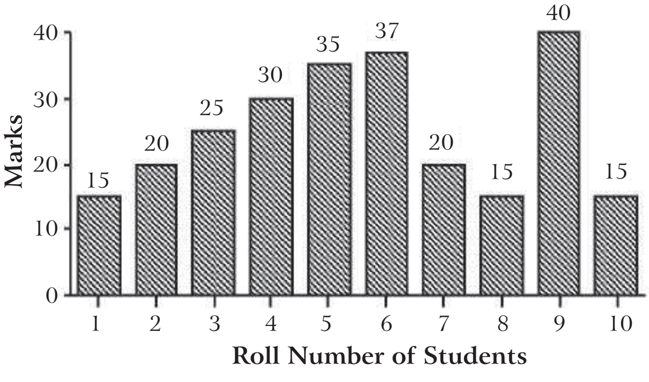


Fig. 3.24: Histogram

Example 22: *Draw a histogram from the following data:*

Marks	1–10	11–20	21–30	31–40	41–50
Students	10	12	20	8	6

Solution: From the data it is clear that it is an inclusive series and therefore, it will be converted into an exclusive series.

Marks	Students
0.5–10.5	10
10.5–20.5	12
20.5–30.5	20
30.5–40.5	8
40.5–50.5	6

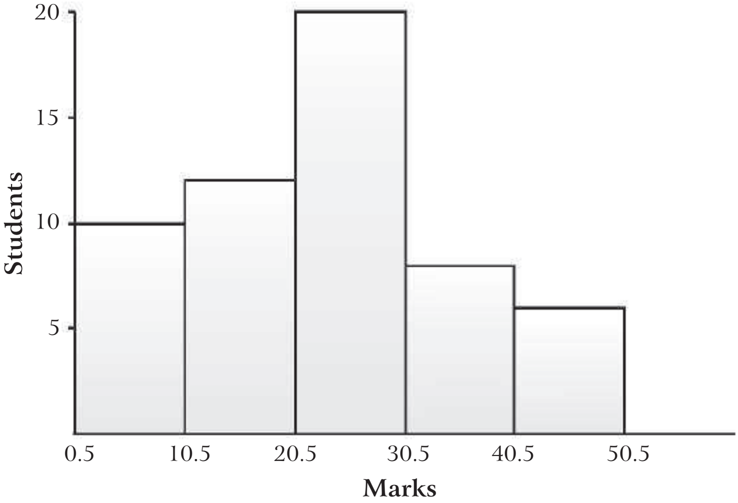


Fig. 3.25: Histogram

Example 23: *Present the following frequency distribution by a histogram:*

Mid-value	2.5	7.5	12.5	17.5	22.5
Frequency	5	10	30	15	10

Solution: In this example mid-values are given and these mid-values are to be converted into class intervals by the method stated below:

Get the difference of second and first mid-values *i.e.,* 7.5 – 2.5 = 5

Divide the difference by 2 *i.e.,* $\left(\frac{5}{2}\right) = 2.5$. The quotient 2.5 is subtracted and added to the first mid-value *i.e.,* 2.5 – 2.5 = 0 and 2.5 + 2.5 = 5. Thus, class interval for first mid value is 0–5. Likewise, other class intervals are obtained for other mid-values, which are as follows:

Class Intervals	0–5	5–10	10–15	15–20	20–25
Frequency	5	10	30	15	10

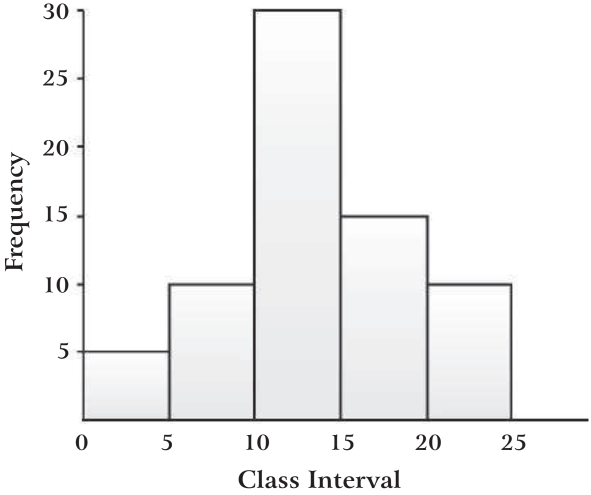


Fig. 3.26: Histogram

Example 24: *Represent the following data by a histogram:*

Weekly Wages (₹)	*Number of Workers*
110–115	6
115–120	18
120–125	25
125–130	15
130–140	12
140–160	12
160–180	8

Solution: The class intervals are unequal and therefore, frequencies are to be adjusted by finding an adjustment factor. Since the lowest class interval is 5, the frequencies of class interval 130–140 shall be divided by 2 because the class interval is double. Similarly, the frequencies of class interval 140–160 and 160–180 shall be divided by 4 each. In other words the adjustment factor for class interval 110–115, 115–120, 120–125 and 125–130 is 1 each, while for 130–140, it is 2 and for 140–160 and 160–180, it is 4 each. This has been summarised in the following table:

Weekly Wage (₹)	Number of Workers	Adjustment Factor 5=1	Frequency
110–115	6	1	6
115–120	18	1	18
120–125	25	1	25
125–130	15	1	15
130–140	12	2	6
140–160	12	4	3
160–180	8	4	2

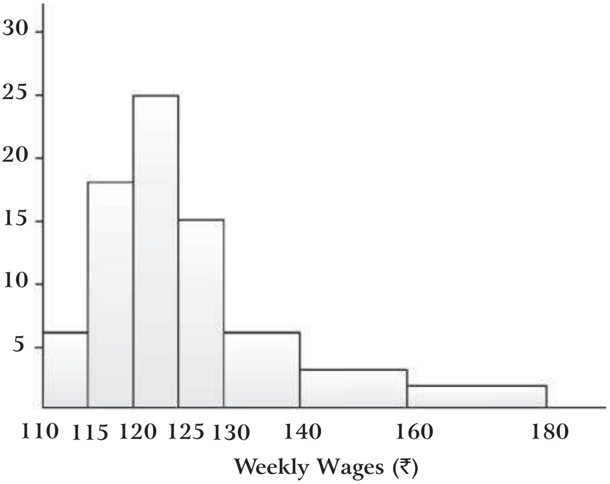


Fig. 3.27: Histogram

Example 25: *Draw a frequency polygon from the following data:*

Classes	0–10	10–20	20–30	30–40	40–50	50–60
Frequency	6	12	16	18	8	4

[B.C.A. (Meerut) 2004]

Solution: We find the mid value of each class and record them as below:

Classes	Mid-value	Frequency
0–10	5	6
10–20	15	12
20–30	25	16
30–40	35	18
40–50	45	8
50–60	55	4

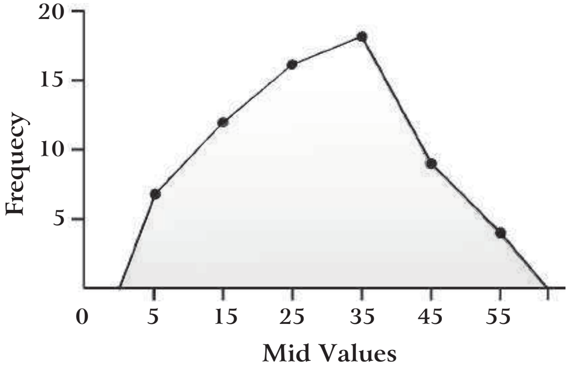


Fig. 3.28: Frequency Polygon

Example 26: *Draw a histogram, frequency polygon and frequency curve from the following data:*

Class Interval	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	4	6	10	20	12	2

[B.C.A. (Meerut) 2010]

Solution:

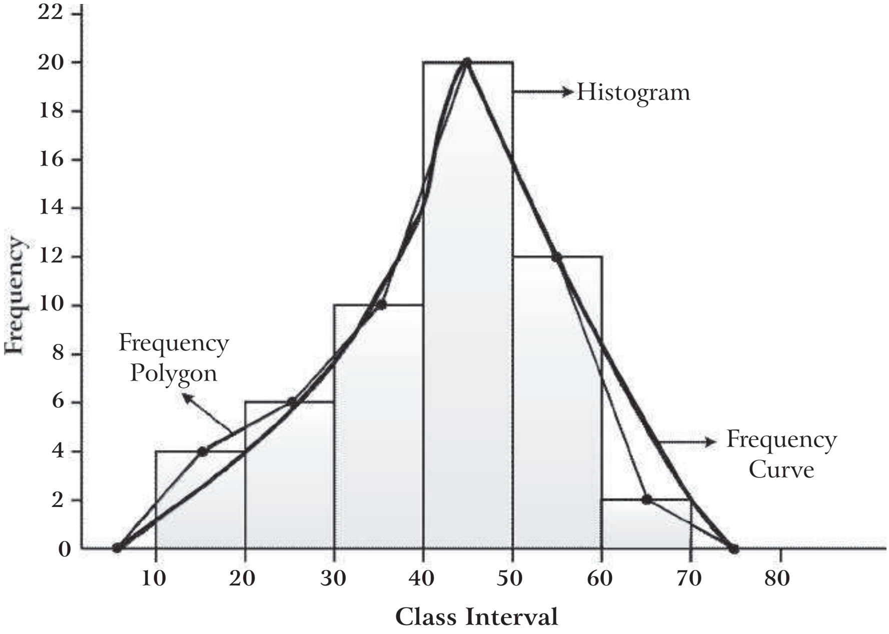


Fig. 3.29: Histogram, Frequency Polygon and Frequency Curve

Example 27:

Wages in (₹)	*No. of Persons*
100–200	3
200–300	10
300–400	14
400–500	24
500–600	17
600–700	14
700–800	3
Total	**85**

Find less than and more than cumulative frequencies and draw ogives. **[B.C.A. (Meerut) 2012]**

Solution: If we cumulate the frequencies of the above distribution, we obtain a cumulative frequency distribution as given below:

Wages in (₹)	No. of Persons	Wages in (₹)	No. of Persons
200	3	100	85
300	13	200	82
400	27	300	72
500	51	400	58
600	68	500	34
700	82	600	17
800	85	700	3

Like a frequency curve of a frequency distribution, we can draw a cumulative frequency curve or ogive of a cumulative frequency distribution. When a series is cumulated upward it is called less than ogive and in this case ogive curve will never fall. Similarly, if a series is cumulated downward, it is called more than ogive and it will be a declining curve.

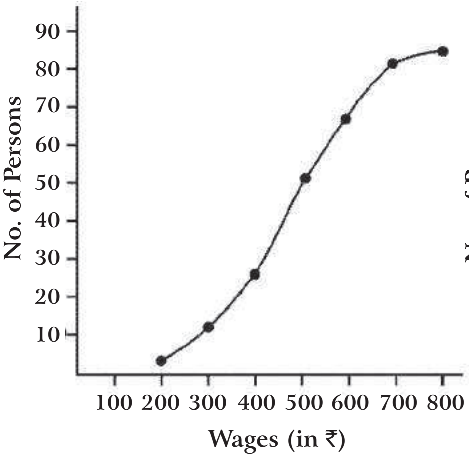
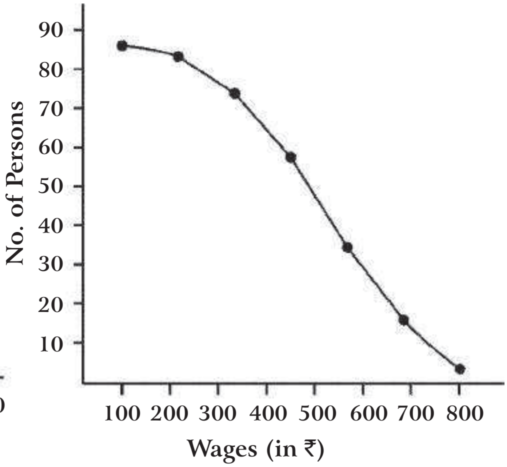


Fig. 3.30: Less than Ogive



3.31: More than Ogive

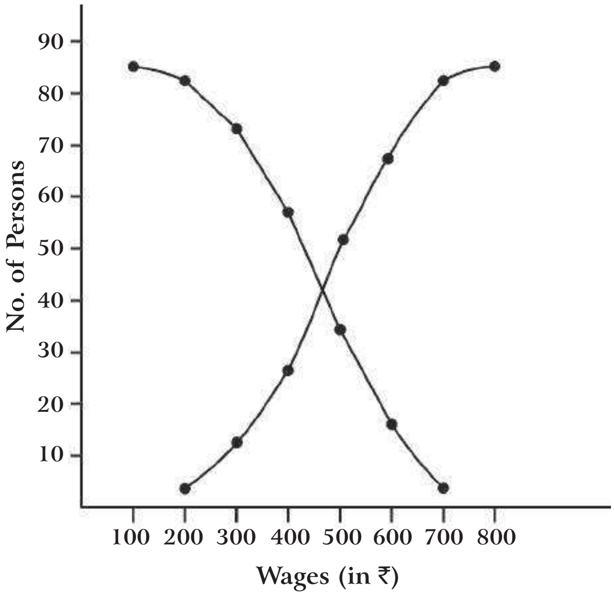


Fig. 3.32: Less than and More than Ogive

Example 28: *Draw histogram, frequency polygon and cumulative frequency curve from the following data:*

Class	10–15	15–20	20–25	25–30	30–35
Frequency	7	19	27	15	12

[B.C.A. (Kanpur) 2006]

Solution:

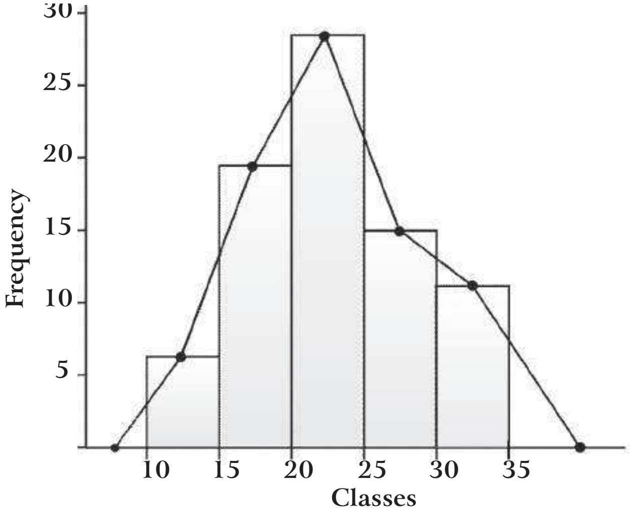


Fig. 3.33: Histogram and Frequency Polygon

For cumulative frequency curve we will find the cumulative frequencies as given below:

Classes (less than)	Frequency	Cumulative Frequency
15	7	7
20	19	26
25	27	53
30	15	68
35	12	80

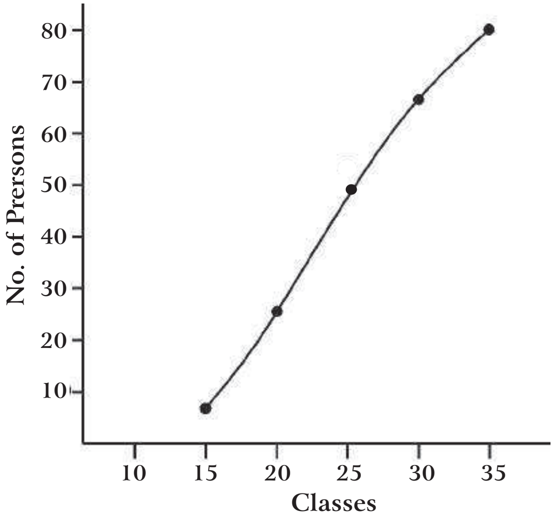


Fig. 3.34: Commutative Frequency Curve (Ogive)

Example 29: *Prepare an ogive from the following data:*

x	*f*
Less than 10	4
Less than 20	8
Less than 30	14
Less than 40	22
Less than 50	28
Less than 60	32
Less than 70	36
Less than 80	38

[B.C.A. (Bundelkhand) 2008]

Solution:

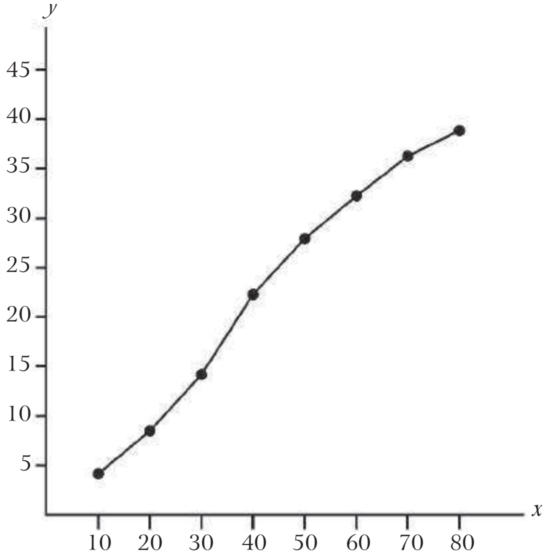


Fig. 3.35: Cumulative Frequency Curve (ogive)

Example 30: *Following is the age distribution of a group of students. Draw the cumulative frequency curve of less than type and hence obtain the median value:*

Age (in Years)	*Frequency*
4 – 5	36
5 – 6	42
6 – 7	52
7 – 8	60
8 – 9	68
9 – 10	84
10 – 11	96
11 – 12	82
12 – 13	66
13 – 14	48
14 – 15	50
15 – 16	16

[B.C.A. (Meerut) 2010]

Solution: From the given table, we may prepare the (less than type) cumulative frequency table as shown on the next page.

Age (in Years)	c.f.
Less than 5	36
Less than 6	78
Less than 7	130
Less than 8	190
Less than 9	258
Less than 10	342
Less than 11	138
Less than 12	520
Less than 13	586
Less than 14	634
Less than 15	684
Less than 16	700

On a graph paper, we take the scale
Along the x-axis: 5 small div. = 1.
Along the y-axis: 1 small div. = 10.

And, plot the points $A(5,36)$, $B(6,78)$, $C(7,130)$, $D(8,190)$, $E(9,258)$, $F(10,342)$, $G(11,438)$, $H(12,520)$, $I(13,586)$, $J(14,634)$, $K(15,684)$, and $L(16,700)$.

We join these points successively with a freehand to get the cumulative frequency curve or an ogive.

Here, $$N = 700 \Rightarrow \frac{N}{2} = 350$$

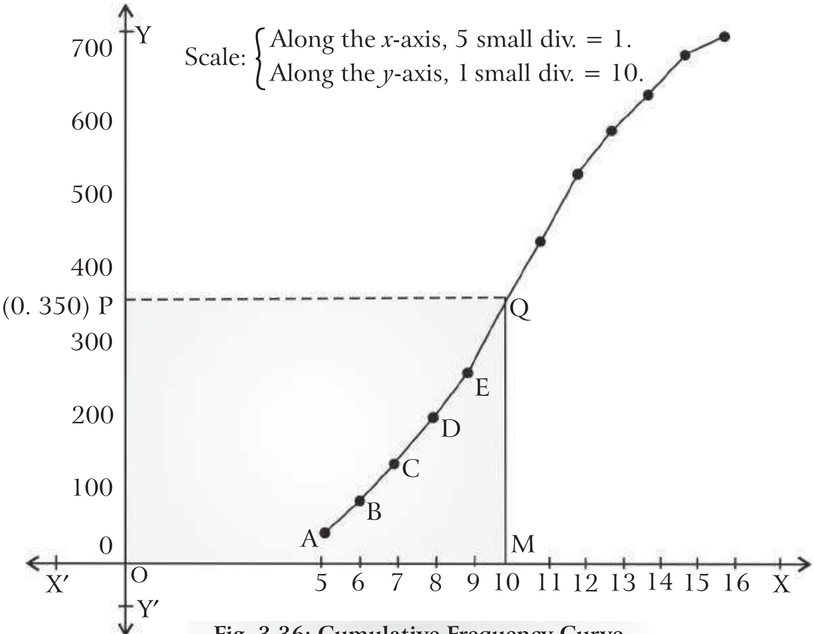


Fig. 3.36: Cumulative Frequency Curve

Take a point $P(0,350)$ on the y-axis and draw $PQ \parallel x$-axis, meeting the curve at Q. Draw $QM \perp$ x-axis, intersecting x-axis at M.

Then, $OM = 10$ units.

Hence, median = 10.

Example 31: *For the following frequency distribution, draw a cumulative frequency curve of more than type and hence obtain the median value:*

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	5	15	20	23	17	11	9

[B.C.A. (Avadh) 2011]

Solution: From the given table, we may prepare the 'more than' series as shown below:

More than 60	9
More than 50	20
More than 40	37
More than 30	60
More than 20	80
More than 10	95
More than 5	100

$$\text{Scale: } \begin{cases} \text{Along the } x\text{-axis, } 10 \text{ small div.} = 5. \\ \text{Along the } y\text{-axis, } 1 \text{ small div.} = 1. \end{cases}$$

Now plot the points $A(5,100)$, $B(10,95)$, $C(20,80)$, $D(30,60)$, $E(40,37)$, $F(50,20)$ and $G(60,9)$.

Joint AB, BC, CD, DE, EF and FG with a freehand to get the required curve, as shown below:

Here, $N = 100 \Rightarrow \frac{N}{2} = 50.$

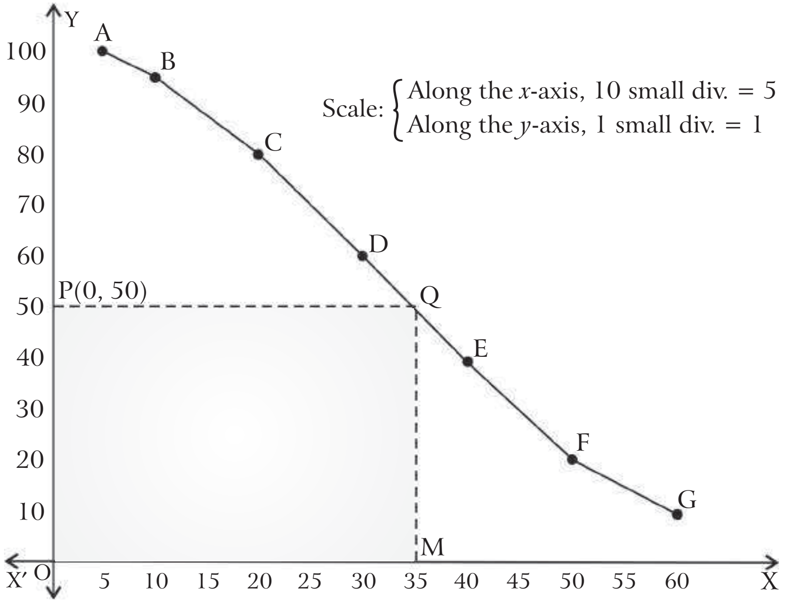


Fig. 3.37: Cumulative Frequency Curve More Than Type

From $P\,(0,50)$ draw $PQ \,||\, x$-axis, meeting the curve at Q. Draw $QM \perp OX$, meeting x-axis at M. Clearly, $OM = 35$ units.

Hence, median = 35.

Example 32: *The following table gives production yield per hectare of wheat of 100 farms of a village:*

Production Yield (kg/hectare)	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65	65 – 70
Number of Farms	4	6	16	20	30	24

Change the distribution to a 'more than type' distribution and draw its ogive.

[M.D.U. (Rohtak) 2011]

Solution: We may prepare the more than series as shown below:

More than 65	24
More than 60	54
More than 55	74
More than 50	90
More than 45	96
More than 40	100

$$\text{Scale:} \begin{cases} \text{Along the } x\text{-axis, 1 small div.} = 1. \\ \text{Along the } y\text{-axis, 1 small div.} = 1. \end{cases}$$

On the graph paper, we plot the points $A\,(40,100)$, $B\,(45,96)$, $C(50,90)$, $D\,(55,74)$, $E(60,54)$ and $F(65,24)$.

Join *AB*, *BC*, *CD*, *DE* and *EF* with a free hand to get a '*More Than Ogive*'.

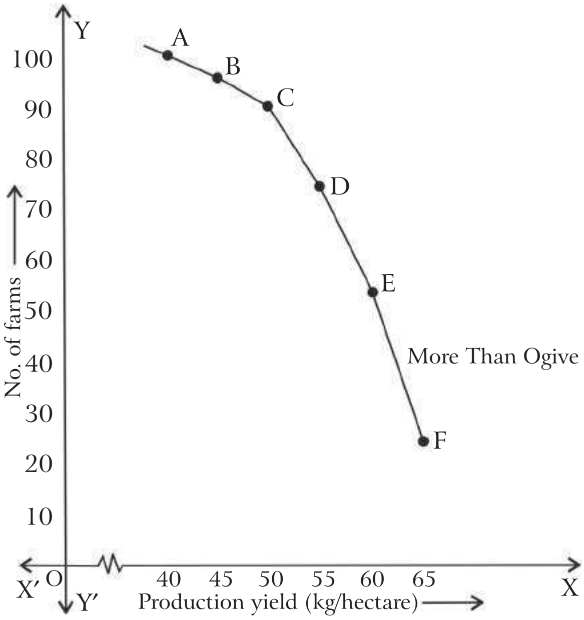


Fig. 3.29: More Than Type Ogive

Example 33: *During the medical checkup of 35 students of a class their weights were recorded as follows:*

Weight (in kg.)	38 – 40	40 – 42	42 – 44	44 – 46	46 – 48	48 – 50	50 – 52
No. of Students	3	2	4	5	14	4	3

Draw a less than type and a more than type ogive from the given data. Hence obtain the median weight from the graph. **[B.C.A (Agra) 2008]**

Solution:

(i) Less than series:

We may prepare the less than series as under:

Weight (in kg.)	Number of Students
Less than 40	3
Less than 42	5
Less than 44	9
Less than 46	14
Less than 48	28
Less than 50	32
Less than 52	35

$$\text{Scale: } \begin{cases} \text{Along the } x\text{-axis}, 5 \text{ small div.} = 1\text{kg.} \\ \text{Along the } y\text{-axis}, 10 \text{ small div.} = 5 \text{ kg.} \end{cases}$$

We plot the points $A(40,3), B(42,5), C(44,9), D(46,14), E(48,28), F(50,32)$ and $G(52,35)$.

Join AB, BC, CD, DE, EF and FG with a free hand to get the curve representing '*less then series*'.

(ii) More than series:

We may prepare the more than series as under:

Weight (in kg.)	Number of Students
More than 38	35
More than 40	32
More than 42	30
More than 44	26
More than 46	21
More than 48	7
More than 50	3

On the same graph-paper as above we plot the points $P(38,35), Q(40,32), R(42,30), S(44,26), T(46,21), U(48,7)$ and $V(50,3)$.

Join PQ, QR, RS, ST, TU and UV with a free hand to get the curve representing '*more than series*'.

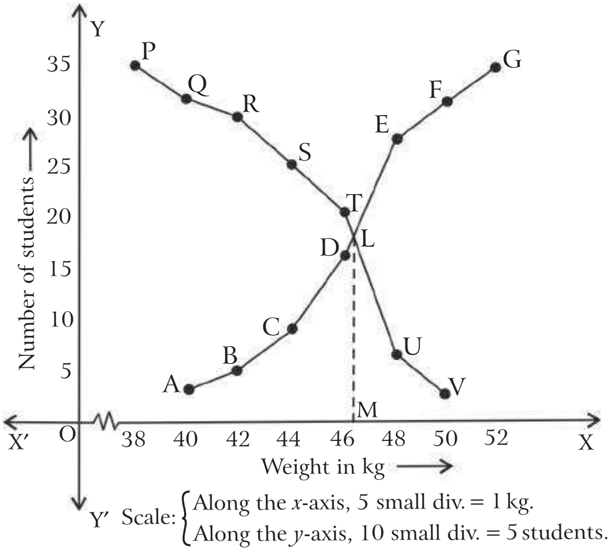


Fig. 3.39: Ogive

The two curves intersect at the point L. Draw $LM \perp OX$.

$\therefore$ Median weight $= OM = 46.5$ kg.

Example 34: *The table given below shows the frequency distribution of the scores obtained by 200 candidates in a B.C.A. entrance examination.*

Score	200–250	250–300	300–350	350–400	400–450	450–500	500–550	550–600
No. of Candidates	30	15	45	20	25	40	10	15

Draw cumulative frequency curves by using (i) 'less than series' and (ii) 'more than series'. Hence, find the median. **[B.C.A. (Lucknow) 2010, 2007]**

Solution:

(i) Less than Series:

We may prepare the less than series as under:

Score	Number of Candidates
Less than 250	30
Less than 300	45
Less than 350	90
Less than 400	110
Less than 450	135
Less than 500	175
Less than 550	185
Less than 600	200

$$\text{Taking the scale:}\begin{cases}\text{Along the x-axis, 1 small div.} = 5.\\ \text{Along the y-axis, 1 small div.} = 2.\end{cases}$$

We plot the point $A\,(250,30)$, $B\,(300,45)$, $C\,(350,90)$, $D\,(400,110)$, $E\,(450,135)$, $F\,(500,175)$, $G\,(550,185)$ and $H\,(600,200)$ on a graph-paper.

Join AB, BC, CD, DE, EF, FG and GH with a free hand to get the curve representing '*less than series*'.

(ii) More than series:

We may prepare the more than series as under:

Score	Number of Candidates
More than 200	200
More than 250	170
More than 300	155
More than 350	110
More than 400	90
More than 450	65
More than 500	25
More than 550	15

Now, on the same graph-paper as above, we plot the points $P(200,200)$, $Q(250,170)$, $R(300,155)$, $S(350,100)$, $T(400,90)$, $U(450,65)$, $V(500,25)$, $G(550,15)$.

Join PQ, QR, RS, ST, TU, UV and VW with a free hand to get the curve representing '*more than series*'.

$$\text{Scale:}\begin{cases}\text{Along the } x\text{-axis}, 1\,\text{small div.} = 5.\\ \text{Along the } y\text{-axis}, 1\,\text{small div.} = 2.\end{cases}$$

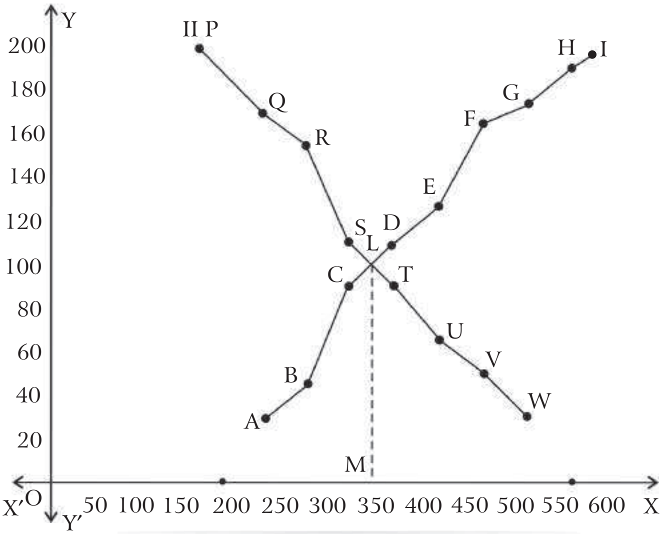


Fig. 3.40: Cumulative Frequency Curves

The two curves intersect at the point L. Draw $LM \perp OX$, cutting x-axis at M.

Clearly, M represents 375.

Hence, median = 375.

Problem Set

1. Explain the method of constructing frequency polygon.
2. An opinion poll in an office showed that 40% were congressites 10% belonged to B.J.P. 41% with communists and 9% none of the above three. Draw a pie diagram.
3. A survey was taken among 100 students of a certain science college to find the most popular subject and the result was shown below:

Subject	Physics	Chemistry	Maths	Statistics	Botany	Zoology
No. of Votes	20	14	21	16	15	14

Present this information by (i) Bar-diagram (ii) Pie-diagram.

[B.C.A. (I.G.N.O.U) 2012]

4. College students of Meerut district are classified according to sex and living condition for 2007-08 as follows:

Particulars	**Boys**	**Girls**	**Total**
College Hostels	10,800	6,100	**16,900**
Private Lodgings	19,000	5,400	**24,400**
Town Residence	19,900	7,100	**27,00**

Construct a bar-diagram to display simultaneously a comparison between the total numbers of students living in each type of accommodation and sex composition of each total.

5. Draw the histogram and frequency polygon for the following data:

x	1	4	7	10	13	16	19	22	25
f	7	45	150	165	180	90	27	18	8

6. Write short notes on the following:

(a) Histogram (b) Frequency polygon

(c) Frequency curve and (d) Ogive

[B.C.A. (Garhwal) 2011; B.C.A. (Bundelkhand) 2010, 2007, 2005]

7. Is there any relation between frequency polygon and frequency curve ?

8. Draw the ogive and frequency polygon for the following distribution:

Class-Interval	0—8	8—16	16—24	24—32	32—40	40—48
Frequency	11	18	20	26	15	10

9. Use rectangle diagram to present the following data:

Items	**Expenditure (₹)**	
	Family A	**Family B**
Food	1500	1200
Clothing	600	500
Rent	800	700
Education	400	300
Miscellaneous	700	800
Total	**4000**	**2500**

10. Draw a cumulative frequency curve (less than type) for the following data and find the median from it: **[B.C.A. (I.P.U., Delhi) 2012; I.G.N.O.U. 2010]**

Class Interval	200–220	220–240	240–260	260–280	280–300	300–320
Frequency	7	3	6	8	2	4

11. Following is the distribution of marks of 70 students in a periodical test:

Marks	Number of Students
Marks less than 10	3
Marks less than 20	11
Marks less than 30	28
Marks less than 40	48
Marks less than 50	70

Draw a cumulative frequency curve for the above data and find the median.

[B.C.A. (Purvanchal) 2005]

12. The following table gives the heights (in meters) of 360 trees:

Height	Number of Trees
Less than 7 m	25
Less than 14 m	45
Less than 21 m	95
Less than 28 m	140
Less than 35 m	235
Less than 42 m	275
Less than 49 m	320
Less than 56 m	360

From the above data, draw an ogive and find the median.

[B.C.A. (Kashi Vidhyapeeth) 2011]

Hint: $\begin{cases} \text{Along the } x\text{–axis, take } 10 \text{ small div.} = 7 \text{ m.} \\ \text{Along the } y\text{–axis, take } 1 \text{ small div.} = 5 \text{ trees.} \end{cases}$

13. From the following frequency distribution, prepare the '*less than ogive*'.

Capital (in Lakh of ₹)	Number of Companies
0 – 10	2
10 – 20	3
20 – 30	7
30 – 40	11
40 – 50	15
50 – 60	7
60 – 70	2
70 – 80	3

Also, find the median. **[B.C.A. (Meerut) 2012]**

14. From the following frequency distribution, prepare the '*more than ogive*':

Score	Number of Candidates
400 – 450	20
450 – 500	35
500 – 550	40
550 – 600	32
600 – 650	24
650 – 700	27
700 – 750	18
750 – 800	34
Total	**230**

Also find the median. **[B.C.A. (Rohilkhand) 2009]**

15. The marks obtained by 100 students of a class in an examination are given below:

Marks	Number of Students
0 – 5	2
5 – 10	5
10 – 15	6
15 – 20	8
20 – 25	10
25 – 30	25
30 – 35	20
35 – 40	18
40 – 45	4
45 – 50	2

Draw cumulative frequency curves by using (i) '*less than series*' and (ii) '*more than series*'.

Hence, find the median. **[B.C.A. (Kashi Vidhyapeeth) 2010]**

16. Use rectangle diagram on percentage basis to represent the data in Problem No. 8.

17. Represent the data in Problem No. 8 by using a multiple Pie-diagram.

18. Explain the method of constructing histogram.

19. Explain the method of constructing ogive.

❑❑❑

Chapter

4

Measure of Central Tendency

4.1 Averages

Objectives of Averages

1. Averages help to get a representative value of the entire mass of data
2. Averages facilitate comparison
3. Averages help in decision making

4.1.1 Characteristics of a Good Average

As the average is a single representative value of the mass of complex data, it must have the following properties:

(i) It should be rigidly defined

(ii) It should be simple to calculate and easy to comprehend

(iii) Its calculation should be based on all the observations

(iv) It should be capable of further algebraic treatment

(v) It should not be affected by extreme values of the observations

(vi) It should be least affected by fluctuations of sampling.

4.2 Measures of Central Tendency

In statistics, various measures of central tendency or averages are studied. Three important measures of central tendency are :

1. Arithmetic Mean (A.M.) or Mean
2. Median
3. Mode

4.2.1 Arithmetic Mean

Arithmetic mean, also called 'average' in general sense, is the most popular and widely used measure of central tendency. Arithmetic mean *is a number obtained by dividing the sum of given observations by their number.* Arithmetic mean is of two types:

1. Simple Arithmetic Mean
2. Weighted Arithmetic Mean

4.2.1.1 Simple Arithmetic Mean

1. **Calculation of Simple Arithmetic Mean:** In classification and tabulation of data, we observed that the values of the variable or observation can be put in the form of any of the following statistical series, namely :
   (i) Individual Series or Ungrouped Data
   (ii) Discrete Frequency Distribution
   (iii) Continuous Frequency Distribution

   (i) Calculation of Arithmetic Mean in Individual Series: Following two methods are used for calculating arithmetic mean of an individual series:
   (a) Direct Method
   (b) Short-cut Method

   (a) Direct Method: The process of calculating arithmetic mean or mean in the case of an individual series is very simple. If the respective values of N observations recorded on a variable X are $X_1, X_2, \ldots, X_N$, then the mean, denoted by $\overline{X}$, is defined by

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{1}{N} \Sigma X \qquad \ldots(1)$$

Here, $\Sigma X = (X_1 + X_2 + \ldots + X_N)$ is the sum of all the N values of the variable X. The formula in (1) can also be written as

$$\Sigma X = N\overline{X} \qquad \ldots(2)$$

It is clear from formula (2) that if any two of the three values of the formula are known, the third one can be calculated.

Solved Examples

Example 1: *Calculate the arithmetic mean of the following marks in statistics obtained by ten students in a class.*

Student No.	1	2	3	4	5	6	7	8	9	10
Marks	15	20	25	19	12	11	13	17	18	20

Solution:

Calculation of Mean using Direct Method

Student No.	Marks (X)
1	15
2	20
3	25
4	19
5	12
6	11
7	13
8	17
9	18
10	20
$N = 10$	$\Sigma X = 170$

Steps:

1. Calculate ΣX by adding all the observations.
2. Divide this total by N, the number of observations. In symbols,

$$\overline{X} = \frac{1}{N} \Sigma X$$

Using formula (1), the mean becomes

$$\overline{X} = \frac{\Sigma X}{N} = \frac{170}{10} = \mathbf{17 \text{ marks}}$$

(b) **Short-Cut Method:** The direct method of calculating arithmetic means is used when the number of items in the series is relatively less. If items are more and figures are large enough, the computation of mean becomes difficult. This difficulty can be solved by using the short-cut method. Under this method an assumed mean is taken as the basis of calculation. *The* assumed mean *is usually chosen to be a neat round number in the middle of the range of the given observations, so that deviations can be easily obtained by subtraction.*

Then, a short-cut formula, based on deviations from assumed mean, for calculating arithmetic mean is:

$$\overline{X} = A + \frac{\Sigma d}{N} \qquad \ldots(3)$$

Here, $\overline{X}$ = the arithmetic mean,

A = the assumed mean,

$d = (X - A)$, the deviation of each value of the variable from the assumed mean A.

Σd = the sum of deviations.

Example 2: *Use short-cut method to calculate arithmetic mean for the data in statistics obtain by ten students.*

Students No.	1	2	3	4	5	6	7	8	9	10
Marks	15	20	25	19	12	11	13	17	18	20

Solution:

Computation of Mean by Short Cut Method

Student No.	**Marks (X)**	**Deviation (X – 18)**
1	15	– 3
2	20	+ 2
3	25	+ 7
4	19	+ 1
5	12	– 6
6	11	– 7
7	13	– 5
8	17	– 1
9	18	0
10	20	+ 2
$N = 10$		$\Sigma d = -10$

In this example, the data ranges from 11 to 25. Therefore, 18, a neat round value in the middle of 11 and 25, may be taken as assumed mean, *i.e.* $A = 18$. The deviations and sum of deviations needed in formula (3) may be calculated from the table given.

Using formula (3), the arithmetic mean will be

$$\overline{X} = A + \frac{\Sigma d}{N}$$

$$= 18 + \left(\frac{-10}{10}\right) = \mathbf{17\ marks.}$$

It can be seen that the arithmetic mean obtained by the direct and the short-cut method is the same.

Example 3: *Show that the arithmetic mean of the series*

$$1, 2, 2^2, 2^3, 2^4, \ldots, 2^n \text{ is } \frac{(2^{n+1} - 1)}{(n+1)}$$

[B.C.A. (Meerut) 2001]

Solution:

$$\text{Required arithmetic mean} = \frac{\text{sum of values of items}}{\text{number of items}}$$

$$= \frac{1 + 2 + 2^2 + 2^3 + \ldots + 2^n}{(n+1)}$$

$$= \frac{\dfrac{(2^{n-1} - 1)}{(2-1)}}{(n+1)}$$

$$= \frac{(2^{n-1} - 1)}{(n+1)}$$

Note: Sum of G.P. of *n*th term

$$S_n = \frac{a(1 - r^n)}{(1 - r)}$$

where a = first term

r = common ratio.

Example 4: *The mean of n numbers of a series is $\bar{x}$ and the sum of first $(n - 1)$ numbers is λ. Find the value of the last number.*

[B.C.A. (Delhi) 2005]

Solution: The sum of the first n numbers

= mean × (number of items)

$= \bar{x} \times n = n\bar{x}$

Also given sum of first $(n - 1)$ numbers = λ

$\therefore$ Required last number = sum of n numbers – sum of first $(n-1)$ numbers

$= n\bar{x} - \lambda$

(ii) Calculation of Arithmetic Mean in a Discrete Frequency Distribution

The following three methods are used for computing arithmetic mean in a discrete frequency distribution:

(a) Direct Method

(b) Short-Cut Method

(c) Step-Deviation Method

(a) Direct Method: In discrete frequency distribution we multiply the values of the variable (X) by their respective frequencies (f) and get the sum of the products (ΣfX). The sum of the products is then divided by the total of the frequencies, *i.e.* $\Sigma f = N$. Thus, according to this method, the formula for calculating arithmetic mean becomes:

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fX}{N} \qquad \ldots(4)$$

Here, ΣfX = the sum of the products of observations with their respective frequencies.

$\Sigma f = N$ = the sum of the frequencies.

Example 5: *Following table gives the wages paid to 125 workers in a factory. Calculate the arithmetic mean of the wages:*

Wage (₹)	200	210	220	230	240	250	260
No. of Workers	5	15	32	42	15	12	4

[B.C.A. (Meerut) 2009]

Solution:

Computation of A.M. (Direct Method)

Wages (X)	No. of Workers (f)	fX
200	5	1000
210	15	3150
220	32	7040
230	42	9660
240	15	3600
250	12	3000
260	4	1040
Total	$N = \Sigma f = 125$	$fX = 28490$

Using the formula (4), the arithmetic mean will be

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{28490}{125} = ₹\mathbf{227.92}$$

(b) **Short-Cut Method:** As observed that short-cut method does not necessarily reduce the computational work in the case of individual series or ungrouped data. However, for grouped data, this method results in considerable saving of time. According to this method, the formula for calculating arithmetic mean is

$$\overline{X} = A + \frac{\Sigma fd}{N} \qquad \text{...(5)}$$

Here, A = the assumed mean,

$d = (X - A)$, the deviation of X from the assumed mean A.

Σfd = the sum of the products of frequencies and their respective deviations.

Example 6: *What is the arithmetic mean of the following ?*

Variate	0	1	2	.... n
Frequency	$^nC_0$	$^nC_1$	$^nC_2$	...... ^nC_n

[B.C.A. (Delhi) 2006, 2001]

Solution:

$$\text{Required mean} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{^nC_0 \,.\, 0 + {^nC_1} \,.\, 1 + {^nC_2} \,.\, 2 + \ldots\ {^nC_n} \,.\, n}{^nC_0 + {^nC_1} + {^nC_2} + \ldots\ {^nC_n}}$$

$$= \frac{n + \dfrac{n(n-1)}{2!} \,.\, 2 + \dfrac{n(n-1)(n-2)}{3!} \,.\, 3 + \ldots 1 \,.\, n}{\text{sum of all binomial coefficient in the expansion } (1+x)^n}$$

$$= \frac{n\left[{^{n-1}C_0} + {^{n-1}C_1} + {^{n-1}C_2} + \ldots + {^{n-1}C_{n-1}}\right]}{2^n}$$

$$= \frac{n.\,[\text{sum of all binomial coefficient in the expansion } (1+x)^{n-1}]}{2^n}$$

$$= \frac{n \,.\, 2^{n-1}}{2^n} = \frac{n}{2}$$

Example 7: *If the arithmetic average of data given below be 165 rupees, find the missing term.*

Monthly Wages in (₹)	100	150	200	—	300	500
Number of Labourers	30	20	15	10	4	1

[B.C.A. (Rohilkhand) 2002]

Solution: Here, taking z as the missing terms we have

Monthly Wages x	No. of Labourers f	fx
100	30	3000
150	20	3000
200	15	3000
z	10	$10z$
300	4	1200
500	1	500
	$\Sigma f = 80$	$\Sigma f(x) = 10700 + 10z$

Now, from Arithmetic average $= \dfrac{\Sigma fx}{\Sigma f}$

$$\Rightarrow \qquad 165 = \frac{10700 + 10z}{80}$$

$$\Rightarrow \qquad 165 \times 80 = 10700 + 10z$$

$$\Rightarrow \qquad 10z = 13200 - 10700$$

$$10z = 2500$$

$$\Rightarrow \qquad z = \frac{2500}{10} = 250$$

Example 8: *Calculate the arithmetic mean for the data in the following table given the wages paid to 125 workers in a factory by using shortcut method:*

Wages (₹)	200	210	220	230	240	250	260
No. of Workers	5	15	32	42	15	12	4

Solution: In this case, assumed mean A is taken as 230 because it is neatly at the centre of the range of the variable X. The other needed calculations are given in the table.

Computation of Arithmetic Mean (Short-Cut Method)

Wages (X)	No. of Workers (I)	Deviations, $d = (X - 230)$	fd
200	5	– 30	– 150
210	15	– 20	– 300
220	32	– 10	– 320
230	42	0	0
240	15	10	+ 150
250	12	20	+ 240
260	4	30	+ 120
	$N = 125$		$\Sigma fd = -260$

Using formula (5), the arithmetic mean will be

$$\overline{X} = A + \frac{\Sigma fd}{N} = 230 + \left(\frac{-260}{125}\right) = 230 - 2.08 = ₹\ \mathbf{227.92}$$

(c) **Step-Deviation Method:** In short-cut method of calculating mean, the deviations taken from an assumed mean generally have a common factor. In continuous frequency distributions this common factor is nothing but the uniform class-interval. The computational work in short-cut method can be further simplified if the deviations obtained are divided by this common factor. The deviations divided by the common factor are called **step-deviations.** According to this method, the arithmetic mean is calculated by the formula:

$$\overline{X} = A + h\frac{\Sigma fd'}{N} \qquad ...(6)$$

Here, h = the common factor in deviations, $d = (X - A)$

$d' = \frac{d}{h} = \frac{X - A}{h}$, the step-deviations.

$\Sigma fd'$ = the sum of the products of the step-deviations and their respective frequencies.

Example 9: *Use step-deviation method for calculating arithmetic mean for the data in following table gives the wages paid to 125 workers in a factory:*

Wage (₹)	200	210	220	230	240	250	260
No. of Workers	5	15	32	42	15	12	4

Solution: While using short-cut method for computing mean in example 8, the deviations (d) have a common factor which is $h = 10$. Thus, the volume of the deviations may be further reduced if we divide these deviations by this common factor. The reduced deviations are called step-deviations and denoted by d'. The other needed calculations are given in the table below:

Computation of Arithmetic Mean (Step-Deviation Method)

Wages X	No. of Workers f	Deviations $d = (X - 230)$	Step-deviations $d' = \frac{d}{10}$	fd'
200	5	– 30	– 3	– 15
210	15	– 20	– 2	– 30
220	32	– 10	– 1	– 32
230	42	0	0	0
240	15	10	1	15
250	12	20	2	24
260	4	30	3	12
	$N = 125$			$\Sigma fd' = -26$

Using formula (6), the arithmetic mean becomes

$$\overline{X} = A + h\frac{\Sigma fd'}{N} = 230 + 10\left(-\frac{26}{125}\right) = 230 - 2.08 = (₹)\,\mathbf{227.92}$$

The value of $\overline{X}$ *obtained by all the three methods is, of course, the same.*

(iii) Calculation of Arithmetic Mean in a Continuous Frequency Distribution: In continuous frequency distribution, the procedure of computing arithmetic mean is the same as in the case of discrete frequency distribution. The only difference is that in continuous frequency distributions the frequencies within each class are assumed to be distributed uniformly over its range. With this assumption, each class is then represented by its mid-point X. *Using these mid-points of the classes and the corresponding frequencies, arithmetic mean for the continuous frequency distribution can be calculated by using any of the methods used in a discrete series.* The following examples will clarify the procedure.

Example 10: *Use direct method to calculate arithmetic mean from the following data:*

Class	20–25	25–30	30–35	35–40	40–45	45–50	50–55
Frequency	8	10	12	20	11	4	5

Solution: For computing arithmetic mean for continuous series, we first obtain mid-points (X) for all the classes. After determining these mid-points, formula (4) is used for computing $\overline{X}$. The calculations needed are shown in the table:

Computation of Mean of a Continuous Series (Direct Method)

Classes	Mid-Points (X)	Frequency (f)	fx
20—25	22.5	8	180.0
25—30	27.5	10	275.0
30—35	32.5	12	390.0
35—40	37.5	20	750.0
40—45	42.5	11	467.5
45—50	47.5	4	190.0
50—55	52.5	5	262.5
Total		$N = 70$	$\Sigma fX = 2515$

Thus, the arithmetic mean will be:

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{2515}{70} = \mathbf{35.93}$$

Example 11: *Use short-cut method for calculating mean for the data given by:*

Class	20—25	25—30	30—35	35—40	40—45	45—50	50—55
Frequency	8	10	12	20	11	4	5

Solution:

Computation of Mean (Short-Cut Method)

Classes	Mid-Points (X)	Frequency (f)	Deviations, $d = X - 37.5$	fd
20—25	22.5	8	– 15	– 120
25—30	27.5	10	– 10	– 100
30—35	32.5	12	– 5	– 60
35—40	37.5	20	0	0
40—45	42.5	11	5	+ 55
45—50	47.5	4	10	+ 40
50—55	52.5	5	15	+ 75
Total		$N = 70$		$\Sigma fd = -110$

Using formula (5), the arithmetic mean can be obtained as

$$\overline{X} = A + \frac{\Sigma fd}{N} = 37.5 + \left(-\frac{110}{70}\right) = 37.5 - 1.57 = \mathbf{35.93}$$

∴ The average is 35.93.

Example 12: *Obtain arithmetic mean for the data in example 2 using step-deviation method.*

Solution:

Calculation of Arithmetic Mean (Step-Deviation Method)

Classes	Mid-Points (X)	Frequency (f)	Deviations, $d = (X - 37.5)$	Step-Deviations $d' = \frac{d}{5}$	fd'
20—25	22.5	8	– 15	– 3	– 24
25—30	27.5	10	– 10	– 2	– 20
30—35	32.5	12	– 5	– 1	– 12
35—40	37.5	20	0	0	0
40—45	42.5	11	5	1	11
45—50	47.5	4	10	2	8
50—55	52.5	5	15	3	15
Total		$N = 70$			$\Sigma fd' = -22$

Here, $A = 37.5$ and $h = 5$ Applying formula (6), the value of $\overline{X}$ will be

$$\overline{X} = A + h\frac{\Sigma fd'}{N} = 37.5 + 5\left(\frac{-22}{70}\right) = 37.5 - 1.57 = \mathbf{35.93}$$

Example 13: *Find the average marks of the students from the following data:*

Marks (below)	10	20	30	40	50	60	70	80
No. of Students	15	35	60	84	96	127	198	250

Solution: In this example a 'less than' cumulative frequency distribution is given and for computing the average, this should first be converted to a simple frequency distribution as shown in the table:

Computation of Mean

Classes	f	X	$d' = \frac{X-35}{10}$	fd'
0—10	15	5	– 3	– 45
10—20	$35-15=20$	15	– 2	– 40
20—30	$60-35=25$	25	– 1	– 25
30—40	$84-60=24$	35	0	0
40—50	$96-84=12$	45	1	12
50—60	$127-96=31$	55	2	62
60—70	$198-127=71$	65	3	213
70—80	$250-198=52$	75	4	208
	$N=250$			$\Sigma fd' = 385$

on using step-deviation formula for computing mean, *i.e.*

$$\overline{X} = A + h\frac{\Sigma fd'}{N} = 35 + 10 \times \frac{385}{250} = 35 + 15.4 = \textbf{50.4 marks}$$

Example 14: *The following data relate to the marks of 70 students in statistics. Find the mean:*

Marks (More than)	20	30	40	50	60	70
No. of Students	70	63	40	30	18	7

Solution: In this example, a 'More than' cumulative frequency distribution is given. For computing mean, the given distribution is converted into a simple frequency distribution as shown in the table.

Computing Mean

Cumulative Classes	Classes	f	X	$d' = \frac{X-55}{10}$	fd'
More than 20	20—30	$70-63=7$	25	-3	-21
More than 30	30—40	$63-40=23$	35	-2	-46
More than 40	40—50	$40-30=10$	45	-1	-10
More than 50	50—60	$30-18=12$	55	0	0
More than 60	60—70	$18-7=11$	65	1	11
More than 70	70—80	$7-0=7$	75	2	14
		$N=70$			$\Sigma fd' = -52$

$$\therefore \quad \overline{X} = A + h\frac{\Sigma fd'}{N} = 55 + 10\left(\frac{-52}{70}\right) = 55 - 7.43 = \mathbf{47.537 \text{ Marks.}}$$

2. **When Inclusive Classes are Given:** Calculation of arithmetic mean remains unaffected even if the frequency distribution with inclusive classes is given. It is because of the fact that the mid values for the inclusive and the corresponding exclusive classes remain the same. The point will be clear from the following example.

Example 15: *Calculate arithmetic mean from the following distribution:*

Classes	10–19	20–29	30–39	40–49	50–59	60–69
Frequency	5	12	17	20	18	8

[B.C.A. (Meerut) 2007]

Solution:

Calculation of Arithmetic Mean

Classes	Mid-Value X	f	$d' = \frac{X-44.5}{10}$	fd'
10–19	14.5	5	-3	-15
20–29	24.5	12	-2	-24
30–39	34.5	17	-1	-17
40–49	44.5	20	0	0
50–59	54.5	18	1	18
60–69	64.5	8	2	16
		$N=80$		$\Sigma fd' = -22$

Here, $A = 44.5,\ h = 10$

$$\therefore \quad \overline{X} = A + h\frac{\Sigma fd'}{N} = 44.5 + 10\left(\frac{-22}{80}\right) = 44.5 - 2.75 = \mathbf{41.75}$$

3. **Mathematical Properties of Arithmetic Mean:** The arithmetic mean has the following important mathematical properties:

   (i) The sum of the deviations of a given set of observations from the arithmetic mean is equal to zero. Symbolically, this property can be written as

$$\Sigma(X - \overline{X}) = 0$$

   Because of this property, mean is called the point of balance *or* the centre of gravity *as the sum of positive deviations from the arithmetic mean is equal to the sum of the negative deviations.*

For Example: ***(i)*** *Here,* $\overline{X} = \frac{\Sigma X}{N} = \frac{200}{5} = 40$

and the sum of deviations from arithmetic mean, i.e.

$$\Sigma(X - \overline{X}) = 0$$

X	$(X - \overline{X})$
20	– 20
30	– 10
40	0
50	10
60	20
$\Sigma X = 200$	$\Sigma(X - \overline{X}) = 0$

   (ii) *The sum of squares of deviations of a set of observations from arithmetic mean is minimum.* In other words, the sum of squared deviations from mean is less than the sum of the squared deviations from any other value. Symbolically,

[B.C.A. (Meerut) 2004, 2003]

$$\Sigma(X - \overline{X})^2 \le \Sigma(X - A)^2$$

   For verification of the property, let us consider an example.

For Example:

X	$(X-\overline{X})$	$(X-\overline{X})^2$	$(X-10)$	$(X-10)^2$
5	– 10	100	– 5	25
10	– 5	25	0	0
15	0	0	+ 5	25
20	+ 5	25	+ 10	100
25	+ 10	100	+ 15	225
$\Sigma X=75$		$\Sigma(X-\overline{X})^2=250$		$\Sigma(X-10)^2=375$

Here, $\overline{X}=\frac{\Sigma X}{N}=\frac{75}{5}=15$

Also, the sum of squared deviations from mean, i.e. $\Sigma(X-\overline{X})^2=250$

Further, let us consider an arbitrary value, say $A=10$*. It is observed from the table that the sum of squared deviations from* $A=10$*, i.e.* $\Sigma(X-A)^2=375$*. Thus,*

$$\Sigma(X-\overline{X})^2=250<\Sigma(X-10)^2=375.$$

(iii) If every value of the variable X is increased (or decreased) by a constant value, the arithmetic mean of the observations so obtained also increases (or decreases) by the same constant.

Symbolically,

Let, $Y=X\pm b$ where b is a constant.

$$\therefore \quad \Sigma Y=\Sigma\,(X\pm b)=\Sigma X+Nb$$

$$\frac{\Sigma Y}{N}=\frac{\Sigma X}{N}\pm b \text{ or } \overline{Y}=\overline{X}\pm b$$

For Example: *Here 5 observations on the variable X are given. A new variable Y is obtained adding 2 (a constant) to each observation on X. Again, another variable Z has been defined by subtracting 3 from each observation on X.*

X	$Y=(X+2)$	$Z=(X-3)$
4	6	1
5	7	2
6	8	3
7	9	4
8	10	5
$\Sigma X=30$	$\Sigma Y=40$	$\Sigma Z=15$

Now,

$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{40}{5} = 8$$

$$\overline{Z} = \frac{\Sigma Z}{N} = \frac{15}{5} = 3$$

On the other hand, using property 3 *above, the values of* $\overline{Y}$ *and* $\overline{Z}$ *can also be obtained as*

$$\overline{Y} = \overline{X} + 2 = 6 + 2 = 8 \text{ (Adding constant 2 to } \overline{X}\text{)}$$

$$\overline{Z} = \overline{X} - 3 = 6 - 3 = 3 \text{ (Subtracting constant 3 from } \overline{X}\text{)}$$

(iv) If the values of the variable X are multiplied (or divided) by a constant, the arithmetic mean of the new observations can be obtained by multiplying (or dividing) the initial arithmetic mean by the same constant. Symbolically,

Let, $Y = kX, k$ being a constant

∴ $\Sigma Y = \Sigma kX = k\Sigma X$

or $$\frac{\Sigma Y}{N} = k\frac{\Sigma X}{N}$$

or $\overline{Y} = k\,\overline{X}$

For Example: *Suppose* 2, 6, 8, 4, 10 *are observations on the variable X. A new variable Y is obtained by multiplying each observation on X by a constant* 3*. Also, another variable Z is obtained by dividing each observation on X by the constant* 2*. Then, the values of observations on Y and Z are shown in the table:*

X	$Y = 3X$	$Z = \frac{X}{2}$
2	6	1
6	18	3
8	24	4
4	12	2
10	30	5
$\Sigma X = 30$	$\Sigma Y = 90$	$\Sigma Z = 15$

Here,

$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6.0$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{90}{5} = 18.0$$

and

$$\overline{Z} = \frac{\Sigma Z}{N} = \frac{15}{5} = 3.0$$

Alternatively, on using property 4 above, $\overline{Y}$ *and* $\overline{Z}$ *can also be obtained as*

$$\overline{Y} = 3\overline{X} = 3 \times 6 = 18 \text{ (on multiplying } \overline{X} \text{ by 3)}$$

$$\overline{Z} = \frac{\overline{X}}{2} = \frac{6}{2} = 3 \text{ (on dividing } \overline{X} \text{ by 2)}$$

(v) If the arithmetic means and the number of observations of two or more related groups are known, we can calculate the combined arithmetic mean of these groups. The combined mean formula for two related groups is as under:

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} \qquad ...(7)$$

Here, $\overline{X}_{12}$ = Combined mean of two group

N_1 = No. of observations in first group

N_2 = No. of observations in second group

$\overline{X}_1$ = mean of the first group

$\overline{X}_2$ = mean of the second group

Similarly, for three related groups, the formula becomes

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3} \qquad ...(8)$$

The formula can also be extended for *k*-groups as

$$\overline{X}_{12...k} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + ... + N_k\overline{X}_k}{N_1 + N_2 + ... + N_k} \qquad ..(9)$$

Here the symbols have their usual meanings.

Example 16: *The mean marks of 60 students in section A is 40 and mean marks of 40 students in section B is 45. Find the combined mean of the 100 students in both the sections.*

Solution: Here, $N_1 = 60, N_2 = 40, \overline{X}_1 = 40$ and $\overline{X}_2 = 45$. Using formula (7), the combined mean of all the 100 students will be

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{60 \times 40 + 40 \times 45}{100} = \frac{2400 + 1800}{100} = \mathbf{42 \text{ marks.}}$$

Example 17: *The mean wage of 100 workers in a factory running two shifts of 60 and 40 workers is 38. The mean wage of 60 workers in the morning shift is ₹ 40. Find the mean wage of 40 workers in the evening shift.*

Solution: Here, $N_1 = 60, N_2 = 40, \overline{X}_1 = 40$ and $\overline{X}_{12} = 38$. We are required to find the value of $\overline{X}_2$. Using the formula (7) for combined mean, *i.e.*

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

We have, $38 = \dfrac{(60 \times 40) + 40\ \overline{X}_2}{100}$

or $3800 - 2400 = 40\ \overline{X}_2$

or $\overline{X}_2 = \dfrac{1400}{40} = ₹\mathbf{35}$

So the mean wage of 40 workers in the evening shift is ₹ 35.

Example 18: *Find the combined mean from the following data:*

Group	1	2	3
Number	200	250	300
Mean	25	20	15

Solution: Here, we are given data related to three groups which can be symbolically put as

$$N_1 = 200, N_2 = 250, N_3 = 300, \overline{X}_1 = 25, \overline{X}_2 = 20, \overline{X}_3 = 15$$

For combined mean, we put these values in formula (8) and get

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3} = \frac{200 \times 25 + 250 \times 20 + 300 \times 15}{200 + 250 + 300}$$

$$= \frac{5000 + 5000 + 4500}{750} = \frac{14500}{750} = \mathbf{19.33}$$

Example 19: *The arithmetic average of a group of* 50 *items is* 12 *and that of another group of* 40 *items is* 14. *Find the arithmetic average of all the* 90 *items.*

Solution: The required mean $= \dfrac{f_1\overline{X}_1 + f_2\overline{X}_2}{f_1 + f_2}$ where f_1, f_2 be frequencies and $\overline{X}_1, \overline{X}_2$ be respective arithmetic average.

$$= \frac{50 \times 12 + 40 \times 14}{50 + 40} = \frac{600 + 560}{90}$$

$$= \frac{1160}{90} = \frac{116}{9} = \mathbf{12.9}$$

Example 20: *The weight of* 150 *students in a certain class is* 60 *kilograms. The mean weight of boys in the class is* 70 *kilograms and that of the girls is* 55 *kilograms find the number of boys and girls in the class.* **[B.C.A. (Meerut) 2004, 2003]**

Solution: The number of students $N = 150$

The mean weight of boys = $\overline{X}_1 = 70$

The mean weight of girls = $\overline{X}_2 = 55$

Let number of boys = N_1

Let number of girls $= N_2$

$\therefore$ $N_1 + N_2 = N \Rightarrow N_1 + N_2 = 150$ ...(1)

Also $\overline{X}_{12} = 60$

We know that

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

$$60 = \frac{70\,N_1 + 55\,N_2}{150}$$

$\Rightarrow$ $70\,N_1 + 55\,N_2 = 9000$

$\Rightarrow$ $14\,N_1 + 11N_2 = 1800$ ...(2)

From (1) $14\,N_1 + 14\,N_2 = 2100$ ...(3)

Subtracting $\Rightarrow$ $3N_2 = 300$

$\Rightarrow$ $N_2 = 100$

and $N_1 = 150 - N_2 = 150 - 100 = 50$

The number of girls $N_2 = 150 - N_1$

$= 150 - 50$

$= \mathbf{100}$.

Example 21: *A person makes a trip which entire travelling* 900 *kms by train at an average speed of* 60 *kph,* 300 *kms by bus at an average speed* 25 *kph,* 400 *kms by plane at* 350 *kph and finally* 15 *kms by taxi* at 25 *kph. What is the average speed of the entire distance.*

[B.C.A. (Meerut) 2006, 2008]

Solution:

Journey by	Distance (km)	Average Speed (kph)	Time Taken (Hrs)
Train	900	60	15
Bus	300	25	12
Plane	400	350	$\frac{8}{7}$
Taxi	15	25	$\frac{3}{5}$
Total	**1615**		$\mathbf{\frac{1006}{35}}$

$$\text{Average Speed} = \frac{\text{distance}}{\text{time}} = \frac{1615}{\frac{1006}{35}} = \frac{1615 \times 35}{1006}$$

$$= \frac{56525}{1006}$$

$$= 56.19 \text{ km/h}$$

And the time can be obtain by

$$\text{Time} = \frac{\text{distance}}{\text{average speed}}$$

Example 22: *A person travels n equal distances with velocities $v_1, v_2 \ldots v_n$. Show that his average velocity $v_1, v_2, v_3, \ldots, v_n$. Show that his average velocity cannot exceed $\frac{(v_1 + v_2 + \ldots v_n)}{n}$. When it will be equal to this value ?* **[B.C.A. (Meerut) 2002]**

Solution: Total distance = n equal distances, velocity per distance v_1, v_2, v_n

$$\text{Average Speed} = \frac{\text{Total distance}}{\text{Total time}}, \quad \text{time} = \frac{\text{distance}}{\text{speed}}$$

$$\text{per unit distance time} = \frac{\text{speed}}{\text{distance}} = \frac{v_1 + v_2 + \ldots + v_n}{n}$$

$$\text{This speed is maximum} \Rightarrow \frac{v_1 + v_2 + \ldots + v_n}{n} \leq \text{Total time}$$

$$\text{Thus, maximum average speed} = \frac{v_1 + v_2 + \ldots + v_n}{n}$$

Example 23: *In a class of 50 students, 10 have failed and their average of marks is 2.5. The total of marks secured by the entire class was 285. Find the average marks of those who have passed.*

Solution: Since the average marks of 10 failed students is 2.5, therefore, the total marks obtained by 10 students who failed $= 10 \times 2.5 = 25$

Now, the marks obtained by all the 50 students=285

∴ the marks obtained by $N = 40$ students who passed $= \Sigma X = 285 - 25 = 260$

∴ the average of marks of these 40 students $= \frac{\Sigma X}{N} = \frac{260}{40} = \mathbf{6.5}$

Example 24: *In the examination, the average grade of all students in class A is 68.4 and that for students in class B is 71.2. If the average of both the class combined is 70, find the number of students in class A to the number in class B.*

Solution: Let the number of students in class A and B be N_1 and N_2 respectively. We are also given $\overline{X}_{12} = 70, \overline{X}_1 = 68.4, \overline{X}_2 = 71.2$. Putting these values in the combined average formula

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

or $$70 = \frac{68.4\, N_1 + 71.2\, N_2}{N_1 + N_2}$$

or $$70\,(N_1 + N_2) = 68.4\, N_1 + 71.2\, N_2$$

Dividing both sides by N_2, one gets

$$70\left(\frac{N_1}{N_2} + 1\right) = 68.4\,\frac{N_1}{N_2} + 71.2$$

$$\frac{N_1}{N_2}(70 - 68.4) = 71.2 - 70$$

or $$1.6\,\frac{N_1}{N_2} = 1.2$$

Thus, the required ratio, *i.e.* $\frac{N_1}{N_2} = \frac{1.2}{1.6} = \frac{3}{4}$

Hence, N_1 and N_2 are in the ratio 3 : 4.

Correcting the Arithmetic Mean

For correcting the incorrect value of mean, first we find the corrected ΣX or ΣfX (in case of discrete or continuous series). For this we subtract the wrong items from the incorrect ΣX (or ΣfX) and add it to the correct items. Finally, on dividing the corrected ΣX by the number of observations, we get the corrected mean.

Example 25: *The average marks of 80 students were found to be 40. Later, it was discovered that a score of 54 was misread as 84. Find the corrected mean of the 80 students.*

Solution: We are given $N = 80,\ \overline{X} = 40$

Since $$\overline{X} = \frac{\Sigma X}{N}$$

$$\therefore \quad \Sigma X = N\overline{X} = 80 \times 40 = 3200$$

But due to the error discovered, $\Sigma X = 3200$ is not correct.

The correct ΣX = incorrect ΣX – misread observation + correct observation.

$$= 3200 - 84 + 54 = 3170.$$

$\therefore$ The corrected average $\overline{X} = \frac{\text{Corrected } \Sigma X}{N} = \frac{3170}{80} = \mathbf{39.625}$

Example 26: *Mean of 100 items is found to be 30. If at the time of calculation, two items are wrongly taken as 32 and 12 instead of 23 and 11, find the correct mean.*

Solution: Given that $N = 100,\ \overline{X} = 30$

$\therefore$ $\Sigma X = N\overline{X} = 100 \times 30 = 3000$ (Incorrect total of 100 items)

Corrected ΣX = Incorrect ΣX – wrong observations + correct observations

$$= 3000 - (32 + 12) + (23 + 11) = 29.90$$

$$\text{Corrected mean} = \frac{\text{Corrected } \Sigma X}{N} = \frac{2990}{100} = \mathbf{29.90}$$

Example 27: *The arithmetic mean of 50 students was given 44. But later on it was found that marks of a student which were read wrongly as '54', there were actually 34. Now calculate the correct mean.*

[B.C.A. (Meerut) 2004]

Solution:

$$\text{The number of students } N = 50$$

$$\text{Incorrect mean } \overline{X} = 44$$

$$\text{Incorrect marks } = 44$$

$$\text{Correct marks } = 34$$

$$\text{Incorrect } \overline{X} = \frac{\text{Incorrect } \Sigma X}{N}$$

$$\text{Incorrect } \Sigma X = 44 \times 50 = 2200$$

$$\text{Correct } \Sigma X = 2200 - 54 + 34 = 2180$$

$$\text{Correct mean } (\overline{X}) = \frac{2180}{50} = \mathbf{43.6}$$

Example 28: *The average monthly wages of all the workers in a factory is ₹ 444. If the average wages paid to male and female workers are ₹ 480 and ₹ 360 respectively, find the percentage of male and female workers employed by the factory.* **[B.C.A. (Meerut) 2008, 2006]**

Solution: Let total workers = x, then $444 = \frac{\Sigma x}{x} \Rightarrow \Sigma x = 444\, x$

$$\text{Again } 480 = \frac{\Sigma x_1}{x}$$

$$\Rightarrow \quad \Sigma x_1 = 480 x_1 \text{ and } \Sigma x_2 = 360\, x_2$$

$$\therefore \quad \Sigma x = \Sigma x_1 + \Sigma x_2$$

where x_1 = number of males, x_2 = number of females

$$\Rightarrow \quad 444x = 480\, x_1 + 360\,(x - x_1)$$

$$\Rightarrow \quad 7x = 10x_1$$

$$\text{Male percentage} = \frac{x_1}{x} \times 100 = \frac{x_1 \times 7}{10\, x_1} \times 100 = 70\%$$

$$\text{Female percentage} = 30\%.$$

Example 29: *Average rainfall of a city from Monday to Saturday is 0.3 inch. Due to heavy rainfall on Sunday, the average rainfall increased to 0.5 inches. What was the rainfall on Sunday ?*

Solution: Given that the average rainfall from Monday to Saturday is 0.3 inch.

Thus, the total rainfall for six days = $N\overline{X} = 6 \times 0.3 = 1.8$ inches

Also given that the average rainfall for all the seven days is 0.5 inch.

∴ Total rainfall for seven days = $7 \times 0.5 =$ **3.5 inches**

∴ The rainfall on Sunday = $3.5 - 1.8 =$ **1.7 inches**

Merits of Arithmetic Mean

The merits and demerits of arithmetic mean can be discussed in the light of characteristics listed for a good average.

(i) It is simple to understand and easy to calculate.

(ii) It is rigidly defined and so its value is always definite.

(iii) Its computation is based on all the observations.

(iv) It is also capable of being handled algebraically, *i.e.* its further algebraic treatment is possible.

(v) Its value is least affected by fluctuations of sampling. In other words, we can say that the arithmetic mean has the minimum sampling variability which is better for drawing inferences about the population.

Demerits of Arithmetic Mean

(i) The value of arithmetic mean is highly affected by extreme values of the variable.

(ii) Arithmetic mean cannot be calculated if even a single observation in the series is missing. Also, in a distribution with open-end classes the mean cannot be determined without making certain assumptions regarding the class-intervals of the open-end classes.

(iii) Arithmetic mean can be a value that does not exist in the series.

4.2.1.2 Weighted Arithmetic Mean

Calculation of Weighted Mean

The formula for calculating weighted arithmetic mean is as follows:

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W} \qquad \text{...(10)}$$

Here,

$\overline{X}_w$ = the weighted arithmetic mean

W = the weights attached to values of the variable

X = the values of the variable

Following few examples will further clarify the uses of weighted mean.

Example 30: *Find the weighted price mean by weighting each price by the quantity consumed for the data given in the table:*

Food Items	Quantity Consumed	Price (₹) (per kg.)
Flour	500 kg.	3.25
Ghee	20 kg.	50.00
Sugar	30 kg.	8.00
Oil	40 kg.	20.00

Solution:

Computation of Weighted Mean

Food Article	Price (₹) X	Quantity (kg.) W	WX
Flour	3.25	500	1625
Ghee	50.00	20	1000
Sugar	8.00	30	240
Oil	20.00	40	800
		$\Sigma W = 590$	$\Sigma WX = 3665$

The weighted mean price will be

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{3665}{590} = ₹\,\mathbf{6.21}$$

1. **Harmonic Mean:** Let $x_1, x_2, \ldots, x_n$ be n values of a variate x. The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the given values *i.e.*

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\Sigma \frac{1}{x_i}}$$

If $f_1, f_2, \ldots, f_n$ are the frequencies of $x_1, x_2, \ldots, x_n$ respectively, then

$$H = \frac{f_1 + f_2 + \ldots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{N}{\Sigma \frac{f_i}{x_i}}$$

If $w_1, w_2, \ldots, w_n$ are the weight of the values $x_1, x_2, \ldots, x_n$ respectively, then

$$H = \frac{w_1 + w_2 + \ldots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots + \frac{w_n}{x_n}} = \frac{N}{\Sigma \frac{w_i}{x_i}}$$

Merits

(i) It is base on all the observations of the series.

(ii) It is capable of the mathematical treatment.

(iii) It is not very much affected by the fluctuations value.

(iv) It is rigidly defined.

Demerits

(i) It is difficult to calculate.

(ii) It is not readily understood.

(iii) It gives higher weight to the smaller values.

2. **Geometric Mean (G.M.):** The geometric mean of 'n' numbers is defined as the "nth root of the product of 'n' numbers". It is found out by multiplying all the values of a series and extracting nth root of the product.

$$\text{Symbolically G.M.} = n\sqrt{(x_1 \,.\, x_2 \,.\, x_3 \ldots x_n)}$$

where, n = number of items.

$x_1, x_2, \ldots, x_n$ be n, non zero observations.

$$\text{or} \qquad \text{G.M.} = \text{Anti log of} \left\{ \frac{\log x_1 + \log x_2 + \ldots + \log x_n}{N} \right\}$$

$$= \frac{\Sigma \log x}{N}$$

Uses of Geometric Mean

(i) Geometric mean is highly useful in averaging ratios percentages and rate of increase between two periods.

(ii) Geometric mean is important in the construction of index numbers.

Merits of Geometric Mean

(i) It is based on all observations.

(ii) It is rigidly defined.

(iii) It is less affected by the extreme values.

(iv) It is useful in studying economic and social data.

Demerits of Geometric Mean

(i) It is difficult to understand.

(ii) Non-mathematical persons cannot do calculations.

(iii) It has restricted application.

Method of Calculation of Geometric Mean

(i) Find out logarithm of each value of the size of item from the log table = log x

(ii) Add all these values of log $x = \Sigma \log x$

(iii) The sum of log ($\Sigma \log x$) is divided by the number of terms $= \dfrac{\Sigma \log x}{N}$

(iv) Find out antilog of this, we get geometric mean of the data.

(v) If $x_1, x_2, \ldots, x_n$ be number of variables with frequencies $f_1, f_2, \ldots, f_n$ then G.M.

$$G = \text{antilog}\left[\frac{1}{N}\sum_{i=1}^{n}(f_i \log x_i)\right]$$

this formula is used for continuous series.

(vi) If G_1 *and* G_2 be geometric means of two distributions in which the number of times are m and n respectively then geometric mean G of the combined distribution is given by

$$\log G = \frac{(m \log G_1 + n \log G_2)}{(m + n)}$$

Example 31: *The daily wages of there persons are* ₹ 6, 9 *and* 13.5. *Find the geometric mean of these wages.*

Solution: If G be the required geometric mean, than

$$G = (6 \times 9 \times 13.5)^{\frac{1}{3}}$$

$$= \left(6 \times 9 \times \frac{27}{2}\right)^{\frac{1}{3}} = (27 \times 27)^{\frac{1}{3}}$$

$$= (3 \times 3 \times 3 \times 3 \times 3 \times 3)^{\frac{1}{3}}$$

$$= 3 \times 3 = ₹ 9$$

Example 32: *Calculate the geometric mean of the following* 2574, 475, 75, 5, 0.8, 0.08, 0.005, 0.0009.

Solution: Here, we have

Value (x)	**Logarithms (log x)**
2574	3.4106
475	2.6767
75	1.8751
5	0.6999
0.8	$\bar{1}.9031$
0.08	$\bar{2}.9031$
0.005	$\bar{3}.6990$
0.0009	$\bar{4}.9545$
	Total = $\Sigma \log x = 2.1220$

Now, if G be the required geometric mean, than

$$\log G = \frac{1}{N} \sum_{i=1}^{N} \log x_i$$

$$= \frac{1}{8} [2.1220] = 0.2652$$

$\therefore$ $$G = \text{Antilog } (0.2652) = 1.841$$

Example 33: *Find the geometric mean for the following data:*

Classes	100–200	200–300	300–400	400–500	500–600
Frequency	15	18	30	20	17

Solution: We form the following table:

Classes	Mid-value (x)	Frequency (f)	$\log_{10} x$	$f \log_{10} x$
100–200	150	15	2.1761	32.6415
200–300	250	18	2.3979	43.1622
300–400	350	30	2.5441	76.3230
400–500	450	20	2.6532	53.0640
500–600	550	17	2.7404	46.5868
		$N = 100$		$\Sigma f \log_{10} x = 251.7775$

Using the formula for geometric mean

$$\log_{10} G = \frac{1}{N} \sum f \log_{10} x$$

$$= \frac{1}{100} (251.7775) = 2.5178$$

$\therefore$ $$G = \text{Antilog } (2.5178) = 329.5$$

Example 34: *Calculate the geometric mean from the following table:*

Marks Obtained	11	12	13	14	15
Number of Students	3	7	8	5	2

Solution: Here, we have

Marks Obtained (x)	Number of Students (f)	log x	f log x
11	3	1.0414	3.1242
12	7	1.0792	7.5544
13	8	1.1139	8.9112
14	5	1.1461	5.7305
15	2	1.1761	2.3522
	$N = 25$		$\Sigma f \log x = 27.6725$

∴ If G be the required geometric mean, then

$$\log G = \frac{1}{N} \Sigma f (\log x)$$

$$= \frac{1}{25} [27.6725]$$

$$\log G = 1.1069$$

$$\Rightarrow \quad G = \text{Antilog } (1.1069) = 12.79$$

Similar Problem: Find the G.M. for the following frequency distribution:

x	2	4	6	8	10	12
f	5	7	15	4	2	1

[B.C.A. (Meerut), 2009]

Example 35: *Find the harmonic mean of* 12, 8, 6 *and* 24.

Solution: Let H be the required harmonic mean.

Than

$$\frac{1}{H} = \frac{1}{4}\left[\frac{1}{12} + \frac{1}{8} + \frac{1}{6} + \frac{1}{24}\right]$$

$$= \frac{1}{4} [0.0833 + 0.1250 + 0.1667 + 0.0417]$$

$$= \frac{1}{4} [0.4167] = 0.1042$$

or $$H = \frac{1}{0.1042} = \frac{10000}{1042} = 9.596$$

Example 36: *Find the harmonic mean from the following table of marks obtained in class test:*

Marks Obtained	11	12	13	14	15
No. of Students	3	7	8	5	2

Solution: Here the computation is better done in tabular form as:

Marks Obtained (*x*)	**Frequency** (*f*)	$\frac{1}{x}$	$\frac{f}{x}$
11	3	0.0909	0.2727
12	7	0.0833	0.5831
13	8	0.07669	0.6152
14	5	0.0714	0.3570
15	2	0.0667	0.1334
	$\Sigma f = 25$		$\Sigma\left(\frac{f}{x}\right) = 1.9614$

∴ If H be the required harmonic mean, we have

$$H = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)} = \frac{25}{1.9614} = \frac{250000}{19614} = 12.746$$

Example 37: *A variate takes values $a, ar, ar^2, \ldots, ar^{n-1}$ each with frequency one. Show that the*

$$A.M. = \frac{a(1-r^n)}{n(1-r)}$$

The $\qquad$ $\text{G.M.} = ar^{\frac{(n-1)}{2}}$ *and the* $H.M. = \frac{a\,.\,n(1-r)\,r^{n-1}}{1-r^n}$.

Prove that $\qquad AH = G^2$

[B.C.A. (Meerut) 2001]

Solution: (i)

$$A = \frac{a + ar + ar^2 + \ldots + ar^{n-1}}{n}$$

$$= \frac{a(1 + r + r^2 + \ldots + r^{n-1})}{n}$$

$$A = \frac{a}{n}\left(\frac{1-r^n}{1-r}\right) \quad \text{(G.P.)}$$

$$G = (a\,.\,ar\,.\,ar^2 \ldots ar^{n-1})^{\frac{1}{n}}$$

$$= [a^n \,.\, r^{1+2+3+\ldots+(n-1)}]^{\frac{1}{n}}$$

$$= a\left[\frac{r^{n(n-1)}}{2}\right]^{\frac{1}{n}} = ar^{\frac{(n-1)}{2}}$$

and $$H = \frac{n}{\frac{1}{a} + \frac{1}{ar} + \frac{1}{ar^2} + \ldots + \frac{1}{ar^{n-1}}}$$

$$= \frac{n}{\frac{1}{a}\left[1 + \frac{1}{r} + \frac{1}{r^2} + \ldots + \frac{1}{r^{n-1}}\right]}$$

$$= \frac{an}{\dfrac{\left(\frac{1}{r^n} - 1\right)}{\left(\frac{1}{r} - 1\right)}} = \frac{an\,(1-r)\,r^{n-1}}{(1-r^n)}$$

Now, $$AH = \frac{a\,(1-r^n)}{n\,(1-r)} \times \frac{an\,(1-r)\,r^{n-1}}{(1-r^n)} = a^2 r^{n-1} = \left[ar^{\frac{(n-1)}{2}}\right]^2 = G^2$$

$\therefore$ $$AH = G^2$$

Example 38: *Find the average rate of (a) motion in the case of a person who rides the first mile at* 10 *m.p.h., the next mile at* 8 *m.p.h. and the third mile at* 6 *m.p.h.; (b) Increase in population which in the first decade had increase* 20%, *in the next* 25% *and in the third* 44%.

[B.C.A. (Delhi) 2007, 2004, 2003]

Solution:

(a) In that case, the harmonic mean being the suitable average, the average rate of motion

$$H = \frac{\text{Total distance}}{\text{Total time taking}} = \frac{3}{\frac{1}{10} + \frac{1}{8} + \frac{1}{6}} = \frac{3 \times 120}{47} = \frac{360}{47} = 7.66 \text{ m.p.p.}$$

(b) Here G.M. is the appropriate average.

$$G = (20 \times 25 \times 44)^{\frac{1}{3}}$$

$$\log G = \frac{1}{3}[\log 20 + \log 25 + \log 44]$$

$$= \frac{1}{3}[1.3010 + 1.3979 + 1.6435]$$

$$= \frac{1}{3} \times 4.3424 = 1.4475 = \log 28.02$$

$\Rightarrow$ $$G = 28.02 = \text{The average increase in population.}$$

Example 39: *A motor when travelling from rest travels the first twentieth of a mile at 6 miles per hour and the next three twentieth of the mile at 8, 12, 24 m.p.h respectively. But its average speed over the first one-fifth of a mile is not 12.5 m.p.h. but 9.6 m.p.h. Explain the apparent paradox.*

[B.C.A. (Rohilkhand) 2004]

Solution: Let A and H are the arithmetic and harmonic means of speeds

$$A = \frac{6+8+12+24}{4} = \frac{50}{4} = 12.5$$

$$H = \frac{4}{\frac{1}{6}+\frac{1}{8}+\frac{1}{12}+\frac{1}{24}} = \frac{4 \times 24}{10} = \frac{48}{5} = 9.6$$

It is obvious from the data that the average speed would not be the arithmetic mean, but it is the harmonic mean, because

$$\text{The average speed} = \frac{\text{Total distance travelled}}{\text{Total time of journey}}$$

$$= \frac{\frac{1}{5}}{\frac{1}{120}+\frac{1}{160}+\frac{1}{240}+\frac{1}{480}}$$

$$= \frac{480}{5 \times 10} = \frac{48}{5} = 9.6 = H.M.$$

Example 40: *A man motors from A to B. A large part of the distance up-hill and he gets a mileage of only 10 miles per gallon of gasoline. On the return trip he makes 15 miles per gallon. Find the harmonic mean of this mileage. Verify that this is the proper average to be used here, assuming that the distance from A to B is 60 miles.*

[B.C.A. (Kanpur) 2007]

Solution: Here, the Harmonic mean,

$$H = \frac{2}{\frac{1}{10}+\frac{1}{15}} = \frac{2 \times 30}{5}$$

$$= 12 \text{ mileage per gallon.}$$

The total distance covered = 120 miles (both ways journey)

$$\text{The average consumption} = \frac{120}{10} = 12 \text{ gallons per mile} = H$$

Hence, harmonic mean is the proper average here.

Example 41: *If the price of a commodity doubles in period of 4 years from* 1997 *to* 2001. *What is the average percentage increase per year.*

Solution: Here $\frac{P_1}{P_0} = 2$ and $n = 4$

$$\therefore \qquad r = \left(\frac{P_1}{P_0}\right)^{\frac{1}{n}} - 1 = (2)^{\frac{1}{4}} - 1 = x - 1$$

$$x = (2)^{\frac{1}{4}}, \log x = \frac{1}{4}\log 2 = \frac{1}{4}(0.3010) = 0.07525$$

$$x = \text{anti} - \log(0.07525) = 1.190$$

$$\therefore \qquad r = 1.190 - 1 = 0.190$$

Average percentage increase $= 0.190 \times 100 = 19$

Example 42: *The geometric mean of a numbers is x and that of b numbers it is y, obtain the G.M. of those a + b numbers combined.*

Solution: $(\text{Product of } a \text{ numbers})^{\frac{1}{a}} = x$

and $(\text{Product of } b \text{ numbers})^{\frac{1}{b}} = y$

$\Rightarrow$ Product of a numbers $= x^a$

and Product of b numbers $= y^b$

The product of $(a + b)$ numbers combined $= x^a y^b$

G.M. of $(a + b)$ numbers $= (x^a y^b)^{\frac{1}{a+b}}$

Example 43: *Find the geometric mean of the series* 1, 2, 4, 8, 16, ... , 2^n

Solution:

$$G = [2^0 . 2^1 . 2^2 . 2^3 \ldots 2^n]^{\frac{1}{(n+1)}}$$

$$= [2^{1+2+3+\ldots n}]^{\frac{1}{(n+1)}}$$

$$= \left[(2)^{\frac{n(n+1)}{2}}\right]^{\frac{1}{(n+1)}} = 2^{\frac{n}{2}}$$

Example 44: *The rise in prices of a certain commodity was* 5% *in* 1994, *8% in* 1995 *and* 77% *in* 1996. *It is said that the average price rise between* 1994 *was* 26% *and net* 30%. *Justify this statement and show that how you would explain it before a layman.*

Solution: The prices in 1994, 1995, 1996 were 105, 108 and 177 respectively, if it was taken as 100 for 1993. Thus, the average of prices for these years would be given by the geometric mean of 105, 108 and 177. Let G be the geometric mean.

Then, $$G = (105 \times 108 \times 177)^{\frac{1}{3}}$$

$$\log G = \frac{1}{3}[\log(105) + \log(108) + \log(177)]$$

$$= \frac{1}{3}[2.0212 + 2.0334 + 2.2480]$$

$$= \frac{1}{3} \times 6.3026 = 2.100\,g$$

$$G = \text{antilog}\,(2.100\,g) = 125.9 = 126$$

Thus the average price rise $= (126 - 100) = 26\%$

If the arithmetic mean n be taken as the average of prices.

Then $$M = \frac{1}{3}[105 + 108 + 177] = 130,$$

Giving the rise $= 30\%$. But if we assume this rise of 30% in every year, the price in 1996

$$= 100 \times \frac{130}{100} \times \frac{130}{100} \times \frac{130}{100} = 219.7$$

Hence the average price rise between 1994 – 1996 is not 30%.

Example 45: *If there are r series of observations $N_1, N_2 \ldots, N_r$, the mean m of the whole series is related to mean $M_1, M_2, \ldots,$ of the component series by the equation.*

$$NM = N_1M_1 + N_2M_2 + \ldots + N_rM_r \text{ where } N = \sum_{r=1}^{n} N_r$$

[B.C.A. (Delhi) 2008; B.C.A. (Kanpur) 2006;
B.C.A. (Rohilkhand) 2002]

Solution: If x_{1j} where $j = 2, 3, 4, \ldots, N_1$ denote the values of the variables in the first series, then the arithmetic mean of this series is given by

$$M_1 = \frac{\sum_{j=1}^{N_1} x_{1j}}{N_1}$$

or $$N_1M_1 = \sum_{j=1}^{N_1} x_{1j} \qquad \ldots(1)$$

Similarly if $x_{2j}, j = 1, 2, 3, \ldots, N_2$ denote the values of the variables in the second series, then the arithmetic mean of this series is given by

$$N_2M_2 = \sum_{j=1}^{N_2} x_{2j} \qquad \ldots(2)$$

Similarly,

$$N_r M_r = \sum_{j=1}^{N_r} x_{rj} \qquad \text{...(3)}$$

Adding (1), (2), (3) we get

$$N_1 M_1 + N_2 M_2 + N_3 M_3 + \ldots + N_r M_r = \sum_{j=1}^{N_1} x_{1j} + \sum_{j=1}^{N} x_{2j} + \ldots + \sum_{j=1}^{N_r} x_{rj}$$

$$= \text{Sum of the values of the variables in the series}$$

$$= \text{NM}$$

Where, M = The mean of the whole series

Theorem 1: The algebraic sum of the deviations of all the variate values from their arithmetic mean is zero.

Proof: Let $x_1, x_2 \ldots., x_n$ be the values of the variables x with corresponding frequency $f_1, f_2, \ldots, f_n$. Then

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} \Rightarrow \Sigma fx = \bar{x}\,\Sigma f \qquad \text{...(1)}$$

Deviation from $\bar{x}$ are given by $d = x - M$, then

$$\Sigma fd = \Sigma f(x - \bar{x}) = \Sigma fx - \bar{x}\,\Sigma f = \bar{x}\,\Sigma f - \bar{x}\,\Sigma f = 0$$

Theorem 2: If every value of the variable is increased by some constant a then arithmetic mean is also increase by a.

Proof: Let $x_1, x_2, \ldots, x_n$ be the values of the variable x with corresponding frequencies $f_1, f_2, \ldots, f_n$. Then $\bar{x} = \frac{\Sigma fx}{\Sigma f}$

Let, $$y = x + a \Rightarrow \bar{y} = \frac{\Sigma fy}{\Sigma f} = \frac{\Sigma f\,(x+a)}{\Sigma f} = \frac{\Sigma fx}{\Sigma f} + \frac{a\,\Sigma f}{\Sigma f}$$

$$= \bar{x} + a$$

4.2.2 Median

In a given 'array' *i.e.* when observations are arranged in an ascending or descending order, every observation holds a certain rank, be that first, second, tenth or forty-second. *One such point which divides the array into two equal parts, so that exactly one-half of the observations are below, and one-half are above the point, is called* **median**. Since the median denotes the central position of an observation in an array, it is called a **position average.** According to Connor, "*The* **median** *is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other values less than the median".*

Determination of Median in Individual Series

The determination of median in an individual series involves the following steps:

(i) The data are arranged either in ascending or descending order.

(ii) Locate median by the value of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item in the order. N being the number of observations in the series.

Example 46: *The following data relate to the height of nine students in a class. Compute the median height:*

S. No.	1	2	3	4	5	6	7	8	9
Heights (in cms)	153	142	151	144	149	146	147	150	154

Solution: For computing median height, we first arrange the data in an ascending order:

142, 144, 146, 147, 149, 150, 151, 153, 154

As a second step, we determine by finding the value of $\frac{N+1}{2} = \frac{9+1}{2} = 5\text{th}$ item in the above order.

∴ **Median** = Size or value of 5th item = **149 cm.**

Example 47: *Calculate median from the following data:*

S. No.	1	2	3	4	5	6	7	8	9	10
Marks	17	32	35	33	15	21	41	32	11	18

Determine the median.

Solution: First we arrange the given marks in ascending order:

S. No.	1	2	3	4	5	6	7	8	9	10
Marks	11	15	17	18	21	32	32	33	35	41

The median is the value of $\left(\frac{N+1}{2}\right)^{th} = \left(\frac{10+1}{2}\right)^{th} = 5.5$ th item.

$$\therefore \qquad \text{Median} = \text{value of 5.5th item} = \frac{\text{value of 5th item + value of 6th item}}{2}$$

$$= \frac{21+32}{2} = \frac{53}{2} = \mathbf{26.5\ marks.}$$

Example 48: *The following table represents the marks obtained by a batch of 20 students in certain class tests in Statistics and Accountancy:*

Serial No.	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	53	54	52	32	30	60	47	46	35	28
Marks (Accountancy)	58	55	25	32	56	85	44	80	33	72
Serial No.	11	12	13	14	15	16	17	18	19	20
Marks (Statistics)	25	42	33	48	72	51	45	33	65	29
Marks (Accountancy)	10	42	15	46	50	64	39	38	30	36

Indicate in which subject is the level of knowledge higher? Give reasons.

Solution: The use of median is recommended when exceptionally large or small values occur in the series. It is also used when one is interested in comparing the average performance of students in different subjects, colleges etc. In view of this median is the most suitable measure of central tendency in the present example. For determining median marks in the two subjects, the marks are first arranged in an ascending order as follows:

Computation of Median

S. No.	Marks in Statistics	Marks in Accountancy	S. No.	Marks in Statistics	Marks in Accountancy
1	25	10	11	46	42
2	28	15	12	47	44
3	29	25	13	48	46
4	30	26	14	51	50
5	32	30	15	52	55
6	33	32	16	53	58
7	33	33	17	54	64
8	35	36	18	60	72
9	42	38	19	65	80
10	45	39	20	72	85

Here, we have 20 observations in both the series. So the median will be the value of

$$\left(\frac{N+1}{2}\right)^{th} = \left(\frac{20+1}{2}\right)^{th} = 10.5\text{th item}$$

$\therefore$ Median = value of 10.5th item = $\dfrac{\text{value of 10th item + value of 11th item}}{2}$

$\therefore$ Median marks (Statistics) = $\dfrac{45+46}{2}=\mathbf{45.5}$

Similarly, Median marks (Accountancy) = $\dfrac{39+42}{2}=\mathbf{40.5}$

The comparison of median marks in the two subjects shows that the level of knowledge is higher in Statistics.

Determination of Median in a Discrete Frequency Distribution

In case of a discrete frequency distribution, the procedure of determining median is fundamentally the same as that for an individual series. The procedure consists of the following steps:

(i) Arrange the data in ascending or descending order or magnitude.

(ii) Obtain the cumulative frequencies.

(iii) Determine the size of $\left(\dfrac{N+1}{2}\right)^{\text{th}}$ item. N being the total frequency.

(iv) Median is located at the value of the variable in whose cumulative frequency the value of $\left(\dfrac{N+1}{2}\right)^{\text{th}}$ item falls.

Example 49: *Find the median size of the following series:*

Size (X)	4	5	6	7	8	9	10
Frequency	6	12	15	28	20	14	5

Solution: As a first step, compute 'less than' cumulative frequencies (C.F.) as shown in the table. Then, for locating median, we find the value of

Computation of Median in Discrete Series

Size X	f	**C.F.**
4	6	6
5	12	18
6	15	33
7	28	61
8	20	81
9	14	95
10	5	100
Total	$N=100$	

$$\left(\frac{N+1}{2}\right)^{th} = \left(\frac{100+1}{2}\right)^{th} = 50.5\text{th}$$

item. Then, as a last step, median is located at the value of the item in whose cumulative frequency the value of 50.5^{th} item falls.

∴ Median = 7

Determination of Median in a Continuous Frequency Distribution

In the case of a continuous frequency distribution, we first locate the median class by cumulating the frequencies until $\left(\frac{N}{2}\right)^{th}$ point is reached. Finally, the median is determined within this class by using an interpolation formula. The procedure thus involves the following steps:

(i) Compute cumulative frequencies.

(ii) Find the size of $\left(\frac{N}{2}\right)^{th}$ item. See that $\left(\frac{N+1}{2}\right)^{th}$ is not used in this case.

(iii) Locate the median class in cumulative frequency column where the size of $\left(\frac{N}{2}\right)^{th}$ item falls.

(iv) Obtain the median value by applying the formula:

$$\text{Median} = l_1 + \frac{\frac{N}{2} - c}{f}(l_2 - l_1) \qquad \text{...(11)}$$

Here, l_1 = lower limit of the median class.

l_2 = upper limit of the median class.

f = frequency of the median class.

c = cumulative frequency of the class preceding the median class.

Example 50: *From the following table find the value of median:*

Measurements	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50
Frequency	7	10	13	26	35	22	11	6

[B.C.A. (Meerut) 2007]

Solution: Here inclusive classes are given. Therefore, for median determination, inclusive class limits need to be converted into class boundaries as shown in the table on the next page.

Calculation of Median (Continuous Frequency Distribution)

Given Classes	Class Boundaries	Frequency (f)	Cumulative Frequency (cf)
11—15	10.5—15.5	7	7
16—20	15.5—20.5	10	17
21—25	20.5—25.5	13	30
26—30	25.5—30.5	26	56
31—35	30.5—35.5	35	91
36—40	35.5—40.5	22	113
41—45	40.5—45.5	11	124
46—50	45.5—50.5	6	130
		$\Sigma f = N = 130$	

Median is the value of $\frac{N}{2} = \frac{130}{2} = 65$th item which lies in the class $(30.5 - 35.5)$. Thus, $(30.5 - 35.5)$ is the median class. For determining the median in this class, we use interpolation formula (11) as under:

$$\text{Median} = l_1 + \frac{\frac{N}{2} - c}{f}(l_2 - l_1)$$

$$= 30.5 + \frac{65 - 56}{35}(35.5 - 30.5)$$

$$= 30.5 + \frac{9 \times 5}{35} = 30.5 + 1.29 = \mathbf{31.79}$$

Example 51: *Calculate the median for the following:*

Marks Below	10	20	30	40	50	60	70	80
No. of Students	15	35	60	84	96	127	198	250

Solution: In such examples, we first convert the given cumulative frequency distribution into the simple frequency distribution as shown in the table on the next page.

Marks Classes	f	$c.f.$
0—10	15	15
10—20	20	35
20—30	25	60
30—40	24	84
40—50	12	96
50—60	31	127
60—70	71	198
70—80	52	250
Total	$N = 250$	

Now, median is the value of $\frac{N}{2} = \frac{250}{2} = 125$th item which lies in (50 – 60). Thus, (50 – 60) is the median class. Using formula (11)

$$\text{Median} = l_1 + \frac{\frac{N}{2} - c}{f}(l_2 - l_1)$$

$$= 50 + \frac{125 - 96}{31} \times 10$$

$$= 50 + \frac{290}{31} = 50 + 9.35$$

$$\mathbf{= 59.35}$$

Example 52: *Following is the distribution of marks of 50 students in a class:*

Marks (more than)	0	10	20	30	40	50
No. of Students	50	46	40	20	10	3

Calculation the median.

Solution: In this example, we are given 'more than' cumulative frequencies. To get the value of median, the given distribution is changed to a simple frequency distribution as shown in the table:

Median is the value of $\frac{N}{2}=\frac{50}{2}=25$th item which lies in (20 – 30) class. For calculating median, we use formula (11) *i.e.*

Classes	f	*c.f.* (less than)
0—10	$50-46=4$	4
10—20	$46-40=6$	10
20—30	$40-20=20$	30
30—40	$20-10=10$	40
40—50	$10-3=7$	47
50—60	$3-0=3$	50
	$N=50$	

$$\text{Median} = I_1 + \frac{\frac{N}{2}-c}{f}(l_2 - l_1)$$

where, $l_1 = 20, l_2 = 30, \frac{N}{2} = 25, f = 20, c = 10$

$$\therefore \quad \text{Median} = 20 + \frac{25-10}{20}(30-20)$$

$$= 20 + \frac{150}{20} = 20 + 7.5 = \mathbf{27.5}$$

Example 53: *Determine median for the following income distribution:*

Income Groups	below 100	100—200	200—300	300—400	400—500	above 500
No. of Persons	5	10	18	30	10	17

Solution: Open-end distribution do not present any difficulty in the location of median as will be clear from the following:

Since median is the value of $\frac{N}{2}=\frac{100}{2}=50$th item which lies in the group $(300-400)$.

Thus $300-400$ is the median class. Finally, the median is determined by using the interpolation formula.

Determination of Median

Groups	f	*c.f.* (less than)
below 100	5	5
100—200	10	15
200—300	18	33
300—400	30	63
400—500	20	83
above 500	17	100
Total	$N = 100$	

$$\text{Median} = l_1 + \frac{\frac{N}{2} - c}{f}(l_2 - l_1)$$

$$= 300 + \frac{50 - 33}{30}(100)$$

$$= 300 + 56.67 = ₹\,\mathbf{356.67}$$

Merits of Median

(i) It is easy to calculate and understand. In some cases it can be located simply by inspection.

(ii) It is a truly defined average as it is the central position of the given data.

(iii) Median can be defined for qualitative data and it is possible to rank the observation in some order.

(iv) It is not affected by extreme observations.

(v) The value of median can also be located graphically.

(vi) Median is specially useful in the case of open-end distributions.

Demerits of Median

(i) For locating median, data have to be arranged in ascending or descending order which is quite tedious for a long series of observations.

(ii) Its determination is not based on all the observations.

(iii) It is not capable of further algebraic treatment.

(iv) Its value is affected by fluctuations of sampling.

(v) In a continuous series, median is calculated by using an intropolation formula. This formula is based on the assumption that all the frequencies of the class are uniformly distributed over the interval which is not always true.

4.2.3 Mode

It is defined as the value of the variate having the maximum frequency.

According to Kenney and Keeping. “The value of the variable which occurs most frequently in a distribution is called the mode”.

The certain cases the maximum frequency may exist in more than one class. In such case there would be more than one mode and the distribution would be bi-modal or multi modal. This generally happens when two a more set of observation are combined in one. For example, wages of skilled and unskilled workers of a factory.

Uses of Mode

(i) It is used when a quick and approximate measure of control tendency is required.

(ii) When the measure of central tendency of the series is one which is most frequently occurring.

Determination of Mode in Individual Series

For an individual series the mode can be located simply by inspection. Here the data can be arranged in an array and then count the frequencies of each variate. The variate having maximum frequency is the mode. Thus, in fact, we always need grouping for mode, since without grouping there would be no frequencies. The following example will clarify the point.

Example 54: *Find the mode of the following array of an individual series of scores:*

9, 9, 10, 11, 11, 11, 12, 14, 14, 17

Solution: Here there are only ten observations in the array, we need not put the data in the form of a frequency distribution and mode can be determined by inspection only. In this case, the observation 11 has the maximum frequency 3. Therefore 11 is the mode.

Mode in Discrete Distribution

In the case of a grouped frequency distribution mode can be located simply by inspection. Here the variate having maximum frequency will be taken as mode.

Example 55: *Determine mode from the following distribution:*

X	10	12	14	16	18	20	22
f	4	6	10	11	21	10	5

Solution: In the above discrete frequency distribution the variate value 18 has the maximum frequency 21. Therefore, 18 is the mode of the given discrete distribution.

Mode by Grouping Method

In a uni-modal distribution where the distribution has only one mode and the highest concentration of values is around one value only, we do not face any difficulty in locating the modal value. However, the difficulty is experienced when nearly equal concentration of the frequencies is observed on two or more neighboring values. In such cases attempt is made to find the value of concentration with the help of grouping method. *In* grouping method *the values are first arranged in ascending order and their frequencies are noted. Normally, the grouping table has the following six columns.*

Column 1 : The maximum frequency is marked by putting a mark or a circle.

Column 2 : The frequencies are grouped in twos and the highest total is marked.

Column 3 : Leaving the first frequency, the remaining frequencies are grouped in twos and the highest total is marked.

Column 4 : The frequencies are grouped in threes. The highest total is marked.

Column 5 : Leaving the first frequency, the remaining frequencies are grouped in threes and the highest total is marked.

Column 6 : Leaving the first two frequencies, the remaining are grouped in threes and the highest total is marked.

After completing the grouping table, an analysis table is formed for finding the value or the observation which is repeated the highest number of times. The same procedure for determining the modal class in the case of a continuous frequency distribution. Let us consider an example for clarifying the procedure.

Example 56: *The following table gives the measurements of collar sizes of* 230 *students in a university. Determine the modal size of the collar:*

Collar-size (cms)	32	33	34	35	36	37	38	39	40	41
No. of Students	7	14	30	28	35	34	16	14	36	16

Solution: Here we observe nearly equal concentration of frequencies for the collar-sizes 36, 37 and 40. Thus, grouping method is used in locating the modal class. Following the steps in grouping method listed above we get the following table on the next page.

Determining Mode by Grouping Method

Collar Size	Frequency					
	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6
32	7	21		51		
33	14		44		72	
34	30	58				93
35	28		63	97		
36	35	69			85	
37	34		50			64
38	16	30		66		
39	14		50		66	
40	36	52				
41	16					

After forming the grouping table, values having highest frequency (shown in bold letters) are counted with the help of an analysis table as shown below:

Analysis Table

Column No.	Values Contributing to the Highest Frequency									
	32	33	34	35	36	37	38	39	40	41
1									✓	
2					✓	✓				
3				✓	✓					
4				✓	✓	✓				
5					✓	✓	✓			
6			✓	✓	✓					
No. of items	—	—	1	3	5	3	1	—	1	—

From analysis table we observe that size 36 is repeated the highest number of times. Therefore, Modal size=36 cms.

Example 57: *Determine the modal value in the following series:*

Value	10	12	14	16	18	20	22	24	26	28	30	32
Frequency	7	15	21	38	34	34	11	19	10	38	5	2

Solution: In this case, equal concentration of values is observed at values 16 and 28. So we determine modal value by grouping method.

Determining Mode by Grouping Method

Value	**Frequency**					
	Col. 1	**Col. 2**	**Col. 3**	**Col. 4**	**Col. 5**	**Col. 6**
10	7	22		43		
12	15		36		74	
14	21	59				93
16	38		72	106		
18	34	68			79	
20	34		45			64
22	11	30		40		
24	19		29		67	
26	10	48				
28	38		43	45		53
30	5	07				
32	2					

Analysis Table

Column No.	**Values Contributing to the Highest Frequency**											
	10	**12**	**14**	**16**	**18**	**20**	**22**	**24**	**26**	**28**	**30**	**32**
1				✓						✓		
2					✓	✓						
3				✓	✓							
4				✓	✓	✓						
5					✓	✓	✓					
6			✓	✓	✓							
Total	—	—	1	4	5	3	1	—	—	1	—	—

Among the highest frequencies in different columns, the value 18 repeats most frequently. Thus, modal value =**18.**

Determination of Mode in a Continuous Series

In a continuous frequency distribution, frequencies are given in various classes. Now a class having maximum frequency is called the modal class. In case nearly equal concentration of frequencies is observed in two or more classes, the grouping method can be used to determine the modal class. After determining the modal class, the precise value of the mode is obtained by using the following interpolation formula.

$$\textbf{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h \qquad ...(15)$$

where,
l_1 = the lower limit of the modal class.
f_1 = the frequency of the modal class.
f_0 = the frequency of the modal class preceding modal class.
f_2 = the frequency of the class succeeding modal class.
h = size of the modal class.

Example 58: *Obtain mode of the following distribution:*

Classes	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	8	12	25	45	11	9

[B.C.A. (Meerut)2007]

Solution: In this example, the highest frequency 45 lies in the class (40 – 50). Thus, (40 – 50) is the modal class. For determining the mode in modal class, we use interpolation formula (15).

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

Here, $f_0 = 25,\ f_1 = 45,\ f_2 = 11,\ l_1 = 40$ and $h = 10$.

$$\text{Mode} = 40 + \frac{45 - 25}{90 - 25 - 11} \times 10 = 40 + 3.70 = \mathbf{43.70.}$$

Example 59: *Locate mode from the following data:*

Mid-value	30	40	50	60	70	80	90
Frequency	7	12	17	29	31	5	3

Solution: In this case, see the nature of the given data. It is not a discrete frequency distribution. Instead, the mid-values of the classes with corresponding frequencies are given. Thus, class-intervals are first determined by using the given mid-values. In this example, maximum concentration of frequencies is at two places, as such, the modal class is determined by grouping method.

Determination of Modal Class by Grouping Method

Mid-Values	Classes	Frequency					
		Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6
30	25–35	7	19		36		
40	35–45	12		29		58	
50	45–55	17	46				77
60	55–65	29		60	65		
70	65–75	31	36			39	
80	75–85	5		8			
90	85–95	3					

Analysis Table

Column No.	Classes Contributing to Highest Frequency						
	25–35	35–45	45–55	55–65	65–75	75–85	85–95
1					✓		
2			✓	✓			
3				✓	✓		
4				✓	✓	✓	
5		✓	✓	✓			
6			✓	✓	✓		
No. of items	—	1	3	5	4	1	—

The analysis table shows that (55 – 65) is the modal class. So by using the interpolation formula (15), *i.e.*

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

Here, $f_0 = 17$

$f_1 = 29$

$$f_2 = 31$$

$$l_1 = 55$$

$$h = 10$$

$$\therefore \qquad \text{Mode} = 55 + \frac{29 - 17}{58 - 17 - 31} \times 10 = 55 + 12 = 67.0$$

But 67.0 does not lie in the modal class (55 – 65) and therefore our answer is not correct. To overcome this difficulty we use interpolation formula (16), *i.e.*

$$\text{Mode} = l_1 + \frac{|f_1 - f_0|}{|f_1 - f_0| + |f_1 - f_2|} \times h$$

$$= 55 + \frac{|29 - 17|}{|29 - 17| + |29 - 31|} \times 10 = 55 + \frac{12 \times 10}{12 + 2}$$

$$= 55 + 12$$

$$= \mathbf{67}$$

Example 60: *Find the mode for the following distribution:*

Class	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	5	10	20	22	24	6	2

Solution: In this example, also, the maximum concentration of frequencies is at three places, as such, the modal class will be determined by grouping method.

Determination of Modal Class by Grouping Method

Classes	**Frequency**					
	Col. 1	**Col. 2**	**Col. 3**	**Col. 4**	**Col. 5**	**Col. 6**
0–10	5	15		35		
10–20	10		30		52	
20–30	20	42				66
30–40	22		46	52		
40–50	24	30			32	
50–60	6		8			
60–70	2					

Analysis Table

Column No.	Classes Contributing to the Highest Frequency						
	0–10	10–20	20–30	30–40	40–50	50–60	60–70
1					✓		
2			✓	✓			
3				✓	✓		
4				✓	✓	✓	
5		✓	✓	✓			
6			✓	✓	✓		
No. of Items	–	1	3	5	4	1	–

The analysis Table shows that (30 – 40) is the modal class. So by using the interpolation formula (15),we have

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

Here, $l_1 = 30,\ f_0 = 20,\ f_1 = 22,\ f_2 = 24$ and $h = 10$

$$\therefore \quad \text{Mode} = 30 + \frac{22 - 20}{44 - 20 - 24} \times 10$$

$$= 30 + \left(\frac{2}{0}\right) \times 10 = \text{infinite}$$

So, in this case also we do not get an admissible value of mode. As such, we apply interpolation formula (16) to get the Modal value, *i.e.*

$$\text{Mode} = l_1 + \frac{|f_1 - f_0|}{|f_1 - f_0| + |f_1 - f_2|} \times h$$

$$= 30 + \frac{|22 - 20|}{|22 - 20| + |22 - 24|} \times 10 = 30 + \frac{2 \times 10}{2 + 2} = 30 + 5$$

$$= \mathbf{35.0}$$

Determining Mode when Class Intervals are Unequal

The formula (15) for locating the mode is applicable where the distribution has equal class-intervals. When the class-intervals are not equal, we must make them equal before calculating the mode. This is done on the assumption that the frequencies are uniformly distributed throughout the class.

Example 61: *Calculate mode from the following table:*

Age (in years)	0–6	6–12	12–18	18–22	22–24	24–26	26–30	30–36	36–42
Frequencies	6	11	25	20	15	8	10	12	6

Solution: In this example, the respective sizes of the class-intervals are 6, 6, 6, 4, 2, 2, 4, 6, 6. However, by properly combining the middle classes of sizes 4, 2, 2 and 4, the series can be amended as under:

The Given Series

Age	Frequency
0–6	6
6–12	11
12–18	25
18–22	20
22–24	15
24–26	8
26–30	10
30–36	12
36–42	6

The Amended Series

Age	Frequency
0–6	6
6–12	11
12–18	25
18–24	35
24–30	18
30–36	12
36–42	6

Hence (18–24) is the modal class in this example. Now the mode for the amended series can be obtained as under

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

Put, $l_1 = 18,\ f_0 = 25,\ f_1 = 25,\ f_2 = 18 \text{ and } h = 6$

we find

$$\text{Mode} = 18 + \frac{35 - 25}{70 - 25 - 18} \times 6 = 18 + \frac{60}{27}$$

$$= 18 + 2.22 = \mathbf{20.22\ years}$$

Example 62: *In a moderately asymmetrical distribution, the mode and mean are* 64.2 *and* 67.4 *respectively. Find the median.*

Solution: The empirical relation between mean, median and mode is :

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

Here, mode = 64.2 and mean = 67.4

∴ $64.2 = 3\ \text{Median} - 2\ (67.4)$

or $$3 \text{ Median} = 64.2 + 134.8 = 199.0$$

or $$\text{Median} = \frac{199}{3} = \mathbf{66.33.}$$

Example 63: *For a symmetrical distribution mean is* 50.6. *Find the values of median and mode.*

Solution: For a symmetrical distribution mean = median = mode

$$\text{Median} = 50.6 \text{ and Mode} = 50.6.$$

Example 64: *For a moderately asymmetrical distribution, the median and mean are* 30.2 *and* 32.4 *respectively. Determine mode.*

Solution: For a moderately asymmetrical distribution,

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

Putting the given values in the above relationship, one gets

$$\text{Mode} = 3\,(30.2) - 2\,(32.4) = 90.6 - 64.8 = 25.8.$$

Merits and Demerits of Mode

Merits

(i) It is comparatively easy to understand.

(ii) It is the simplest descriptive measure of average.

(iii) It is not affected by extreme items. It can be obtained even if the extreme values are not given.

(iv) It can be determined for open-end distributions.

(v) It can be located graphically while mean cannot be ascertained through graphs.

(vi) Mode has been defined as the most typical value of a distribution. Therefore, it is a useful average for many practical situations, such as, average size of shoe, average price of a commodity, the average type of dress, average wages and so on.

Demerits

(i) It is not precisely defined.

(ii) It is not based on all the observations.

(iii) It is not capable of being handled algebraically as its value is not based on all the observations.

(iv) The mode does not exist in many cases while there may be more than one mode in other cases, *i.e.* it is not useful as an average in such situations.

(v) The value of mode is significantly affected by the size of the class-interval which is the basis of grouping the frequencies.

Example 65: *A distribution $x_1, x_2, \dots, x_n$ with frequencies $f_1, f_2, \dots, f_n$ is transformed into the distribution $X_1, X_2, \dots, X_n$ with the same corresponding frequencies of the relation*

$$X_i = ax_i + b\,, \quad i = 1, 2, 3, \dots, n$$

where a and b are constants. Show that the mean, mode and median of the new distribution are given in terms of those of the first distribution by the same transformation.

[B.C.A. (Delhi) 2006; B.C.A. (Kanpur) 2004; B.C.A. (Meerut) 2002]

Solution: Given $X_i = ax_i + b\,,\; i = 1, 2, 3, \dots, n$

(i) Let $\overline{X}$ and $\bar{x}$ be the means of the X_i and x_i values respectively. Then

$$\overline{X} = \frac{\sum_i f_i X_i}{\sum_i f_i} = \frac{\sum_i f_i (ax_i + b)}{\sum_i f_i} = \frac{a \sum_i f_i x_i}{\sum_i f_i} + b$$

$$= a\bar{x} + b$$

Thus, the arithmetic mean is given by the same transformation.

(ii) $$M_d = L_0 + \frac{\frac{N}{2} - F}{l} X_i\,, \quad X_i = ax_i + b$$

It will make to the lower limit of X_i, change into $aL_0 + b$ and interval i will change to a_i.

Thus, new median $(M_d') = (aL_0 + b) + \dfrac{\frac{N}{2} - F}{f} a_i$

$$= a\left\{L_0 + \frac{\frac{N}{2} - F}{f} i\right\} + b$$

$$= a\,M_d + b$$

(iii) $$M_0 = L_o + \frac{f_1 - f_{-1}}{2\,f - f_{-1} - f_{+1}} X_i, \quad X_i = ax_i + b$$

Lower limit will become $ax_i + b$ and interval $= a_i$

The new $\quad M_0' = (aL_0 + b) + \dfrac{f_1 + f_{-1}}{2 f_1 - f_{-1} - f_{+1}}$

$$= a\left\{L_0 + \frac{f_1 + f_{-1}}{2 f_1 - f_{-1} + f_{+1}}\right\} + b = aM_0 + b$$

Example 66: *In what respect A.M. is superior to mode and median as measure of central tendency.*

[B.C.A. (Meerut) 2003, 2002]

Solution:

(i) A.M. is capable of further mathematical treatment whereas mode and median have no such property.

(ii) A combined mean of many series can be obtained, where as evaluation of combined mode and median through some formula is impossible.

(iii) Arithmetic mean is least affected by the fluctuations of sampling, whereas mode and median are in victim of it.

(iv) Median and mode can have end values of the series but A.M. will have the central value. Thus, median and mode may not be true representative of the data.

(v) A.M. depends upon all the observations, but medium and mode do not.

(vi) A.M. is very well defined. But median and mode both are comparatively ill defined measures.

Example 67: *The monthly incomes of 10 families are given by the following table:*

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*Total*
85	70	15	75	500	20	45	250	40	36	1136

Compute the arithmetic mean, geometric mean and harmonic mean of the above incomes. Which of above three means represents the data most suitable? Give reasons.

Solution: Here arithmetic mean

$$A = \frac{\Sigma x_i}{n} = \frac{1136}{10} = ₹\, 113.6$$

Geometric mean

$$G = (85 \times 70 \times 15 \times 75 \times 500 \times 20 \times 45 \times 250 \times 40 \times 36)^{\frac{1}{10}}$$

$$\Rightarrow \qquad \log G = \frac{1}{10}[\log 85 + \log 70 + \log 15 + \log 75 + \log 500$$

$$+ \log 20 + \log 45 + \log 250 + \log 40 + \log 36]$$

$$= \frac{1}{10}[1.9294 + 1.8451 + 1.1761 + 1.8751 + 2.6980 + 1.3010$$

$$+ 1.6532 + 2.3979 + 1.6021 + 1.55631]$$

$$= \frac{1}{10} \times 18.0342 = 1.8034 = \log 63.59$$

$$\Rightarrow \qquad G = ₹\, 63.59$$

And harmonic mean.

$$H = \frac{10}{\frac{1}{85} + \frac{1}{70} + \frac{1}{15} + \frac{1}{75} + \frac{1}{500} + \frac{1}{20} + \frac{1}{45} + \frac{1}{250} + \frac{1}{40} + \frac{1}{36}}$$

$$= \frac{10}{[0.01176 + 0.0143 + 0.0667 + 0.0133 + 0.002 + 0.05 + 0.0222 + 0.004 + 0.025 + 0.0278]}$$

$$= \frac{10}{0.23706} = 42.18$$

Clearly the geometric mean represents the data most suitable since out of 10 families, 8 families have their income less than the arithmetic mean while the harmonic mean represents the lower income group only.

Example 68: *The expenditure of 100 families are given by:*

Expenses in (₹)	0–10	10–20	20–30	30–40	40–50
No. of Families	14	?	27	?	15

The mode for the distribution is 24. Find the missing frequencies.

Solution: Let F_1 and F_2 be the unknown frequencies of class 10–20 and 30–40 respectively.

Expenses in (₹)	No. of Families (f)	Commutative Frequency
0–10	14	14
10–20	F_1	$14 + F_1$
20–30	27	$41 + F_1$
30–40	F_2	$41 + F_1 + F_2$
40–50	15	$56 + F_1 + F_2$
	$N = \Sigma f = 56 + F_1 + F_2$	

$\therefore$ $N = 100 = 56 + F_1 + F_2$

$\Rightarrow$ $F_1 + F_2 = 44$ ...(1)

Here mode is 24 which clearly lies in the class 20–30 and so 20–30 is the modal class. Therefore from the formula, made M_0 is given by

$$M_0 = l + \frac{f - f_{(-1)}}{2f - f_{-1} - f_1} \times h$$

where, $l = 20, f = 27, f_{(-1)} = F_1, f_1 = F_2$ and $h = 10$

$\therefore$ $$24 = 20 + \frac{27 - F_1}{54 - F_1 - F_2} \times 10$$

$\Rightarrow$ $3F_1 - 2F_2 = 27$ ...(2)

Solving (1) and (2), we have

$F_1 = 23, F_2 = 21$

Example 69: *A man travelled by car for 3 days. He covered 480 kms each day. On the first day he draw for 10 hours at 48 kms an hour, on the second day the draw for 12 hours at 40 kms an hour and on last day he draw for 15 hours at 32 kms. What was his average speed.*

[B.C.A. (Rohilkhand) 2010, 2008]

Solution: Since the distance travelled is constant *i.e.*, 480 kms each day, the appropriate average is the Harmonic mean.

$$\text{H.M.} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}}$$

Here $N = 3, x_1 = 48, x_2 = 40, x_3 = 32$

$$\Rightarrow \qquad \text{H.M.} = \frac{3}{\frac{1}{48} + \frac{1}{40} + \frac{1}{32}} = \frac{3 \times 480}{37} = 38.92$$

Example 70: *The following tables shows certain data collected for the three regions of a country:*

Region	*No. of Inhabitants (In million)*	*Percent of Literates*	*Average Annual Income per person (₹)*
A	10	52	8000
B	5	68	5800
C	20	41	7500

Obtain the above figures for the three regions taken together.

[B.C.A. (Rohtak) 2009; B.C.A. (Rohilkhand) 2006]

Solution: Total population $= 10 + 5 + 20 = 35$ million

Average population $= 35/3 = 11.6$ million

$$\text{Total number of literates} = 10 \times \frac{52}{100} + 5 \times \frac{68}{100} + 20 \times \frac{41}{100} = 5.2 + 3.4 + 8.2$$

$$= 1.6 \text{ millions.}$$

$$\text{Percentage of literacy in overall region} = \frac{16.8}{35} \times 100 = 48\%$$

$$\text{Overall average annual income person} = \bar{x}_{123} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3}{x_1 + x_2 + x_3}$$

$$= \frac{10 \times 8000 + 5 \times 5800 + 20 \times 7500}{10 + 5 + 20} = \frac{259000}{35}$$

$$= ₹\ 7400$$

Example 71: *Calculate the values of Arithmetic mean, mode and standard deviation for the following distribution:*

Wage (₹)	*No. of Workers*
0–10	5
10–20	12
20–30	20
30–40	32
40–50	60
50–60	82
60–70	92
70–80	100

[B.C.A. (Meerut) 2011, 2009; B.C.A. (Delhi) 2011, 2004, 2002]

Solution:

Class Interval ($C.I$)	f	Mid Value (m)	$d=(x-A)$ $A=35$	$d'=\frac{d}{10}$	fd'	fd'^2
0–10	5	5	−30	−3	−15	45
10–20	7	15	−20	−2	−14	28
20–30	8	25	−10	−1	−8	8
30–40	12	35	0	0	0	0
40–50	28	45	10	1	28	28
50–60	22	55	20	2	44	88
60–70	10	65	30	3	30	90
70–80	8	75	40	4	32	128
Total	$N=100$				$\Sigma fd'=97$	$\Sigma fd'^2=415$

$$\text{A.M.} = \overline{X} = A + \frac{\Sigma fd'}{N} \times i, \text{ Here } A = 35, \Sigma fd' = 97, N = 100$$

$$i = 10 = \text{length of class interval}$$

$$\Rightarrow \qquad \overline{X} = 35 + \frac{97}{100} \times 10 = 35 + 9.7 = 44.7$$

$$\therefore \qquad \text{Mean} = 44.7$$

By inspection, modal class interval is 40–50

$$\text{Mode }(Z) = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $l_1 = 40,\ f_1 = 28,\ f_0 = 12,\ f_2 = 22,\ i = 10$

$$\Rightarrow \quad \text{Mode }(Z) = 40 + \frac{28 - 12}{2 \times 28 - 12 - 22} \times 10$$

$$= 40 + \frac{16}{22} \times 10 = 40 + 7.27 = 47.27$$

$$\therefore \quad \text{Mode }(Z) = 47.27$$

$$\text{Standard deviation }(\sigma) = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma fd'}{\text{N}}\right)^2} \times i$$

Here $\Sigma fd'^2 = 415,\ \Sigma fd' = 97,\ i = 10,\ N = 100$

$$\therefore \quad \sigma = \sqrt{\frac{415}{100} - \left(\frac{97}{100}\right)^2} \times 10$$

$$= \sqrt{4.15 - 0.94} \times 10 = \sqrt{3.21} \times 10$$

$$= 1.792 \times 10 = 17.92$$

$\therefore$ Standard deviation $= 17.92$

Example 72: *In what respect A.M. is superior to made and median as a measure of central tendency ?*

[B.C.A. (M.D.U) Rohtak 2011; B.C.A. (Agra) 2010, 2007, 2005; B.C.A. (Meerut) 2008; B.C.A. (Rohilkhand) 2008]

Solution:

(i) A.M. is very well defined. But median and mode both are comparatively ill defined measures.

(ii) A.M. depends upon all the observations, but median and mode do not.

(iii) A.M. is capable of further mathematical treatment whereas mode and median have no such property.

(iv) A combined mean of many series can be obtained, whereas evaluation of combined mode and median through some formula is impossible.

(v) Median and mode can have end values of the series but A.M. will have the central value. Thus median and mode may not be true representative of the data.

(vi) Arithmetic mean is least affected by fluctuations of sampling, whereas mode and median are in the victim of it.

Example 73: *Give an idea of the uses and limitations of averages in summarising a given data.*
[B.C.A. (Kurukshetra) 2011; B.C.A. (Lucknow) 2009, 2005]

Solution:

(i) Averages are used for obtaining variation, skewness, kurtosis and many other basic characteristics of a data.

(ii) Averages help us in considering the mass of the data in one single value which represents the entire group.

- **Limitations of Averages**
  (i) The averages do not tell anything about the nature and composition of the data under study. So the averages should not be expected to tell more than what they can.
  (ii) An average in itself is single simple value, so it can yield wrong interpretations too.
  (iii) At times, average in themselves fail to convey about the nature of the series, because two quite different series may have the same average.

Example 74: *Evaluate the values of mean, mode and median for the following grouped cumulative data:*

No. of Days Absent	*No. of Students (f)*
Less than 5	29
Less than 10	124
Less than 15	349
Less than 20	442
Less than 25	478
Less than 30	487
Less than 35	493
Less than 40	497
Less than 50	500

[B.C.A. (Lucknow) 2009, 2006; B.C.A. (Bundelkhand) 2008]

Solution: These results can be tabulated as given below:

No. of Days Absent (class)	Mid Value (x)	No. of Students Frequency (f)	Cumulative Frequency	Deviation from Assumed Mean A (22.5), $d = x - A$	fd
0–5	2.5	29	29	– 20	–580
5–10	7.5	95	124	–15	– 1425
10–15	12.5	225	349	–10	–2250
15–20	17.5	93	442	– 5	– 465
20–25	22.5	36	478	0	0
25–30	27.5	9	487	5	45
30–35	32.5	6	493	10	60
35–40	37.5	4	497	15	60
40–50	45.0	3	500	22.5	67.5
	Total	$\Sigma f = N = 500$			$-4488.5 = \Sigma fd$

$\therefore$ $$\text{Mean} = A + \frac{\Sigma fd}{N} = 22.5 + \frac{-4488.5}{500} = 22.5 - 8.977 = 13.5$$

$$\text{Median} = l_2 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Here, $l_2 = 10,\ c.f. = 124,\ f = 225,\ i = 5,\ N = 500$

$\therefore$ $$\text{Median} = 10 + \frac{250 - 124}{225} \times 5 = 10 + \frac{630}{225} = 10 + 2.8 = 12.8$$

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times l = 10 + \frac{225 - 95}{2 \times 225 - 95 - 93} \times 5$$

$$= 10 + \frac{130 \times 5}{450 - 188} = 12.48.$$

Example 75: *The mean of the following frequency distribution is 62.8 and the sum of all frequencies is 50. Compute the missing frequency f_1 and f_2.*

Class	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 110	Total
Frequency	5	f_1	10	f_2	7	8	50

[B.C.A. (Agra) 2009; B.C.A. (Meerut) 2006]

Solution: We have

$$5+f_1+10+f_2+7+8=50 \Rightarrow f_2=20-f_1$$

Now, we may prepare the table given below:

Class Interval	**Frequency (f_i)**	**Mid Value (x_i)**	**($f_i x_i$)**
0 – 20	5	10	50
20 – 40	f_1	30	$30f_1$
40 – 60	10	50	500
60 – 80	$20-f_1$	70	$1400-70f_1$
80 – 100	7	90	630
100 – 120	8	110	880
Total	$\Sigma f_i=50$		$\Sigma f_i x_i=(3460-40f_1)$

$$\therefore \qquad \text{Mean}=\frac{\Sigma f_i x_i}{\Sigma f_i}=\frac{3460-40f_i}{50}, \text{But mean} = 62.8$$

$$\therefore \qquad \frac{3460-40f_1}{50}=62.8 \Rightarrow 3460-40f_1=3140$$

$$\Rightarrow 40f_1=320 \text{ or } f_1=8$$

This, $\qquad f_1=8$

and $\qquad f_2=(20-8)=12$

Remark: We can use the formula for calculating the mode as:

Mode, $$M_o=x_k+h.\left\{\frac{(f_k-f_{k-1})}{(2f_k-f_{k-1}-f_{k+1})}\right\},$$

Where,

x_k = lower limit of the modal class interval

h = width of the class interval

f_k = frequency of the modal class

f_{k-1} = frequency of class preceding the modal class

f_{k+1} = frequency of class succeeding the modal class.

Example 76: *Calculate the mode for the following frequency distribution:*

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	5	8	7	12	28	20	10	10

[B.C.A. (Meerut) 2010]

Solution: As the class 40 – 50 has the maximum frequency, so it is the modal class.

∴ $x_k = 40, h = 10, f_k = 28, f_{k-1} = 12$ and $f_{k+1} = 20$

Mode, $$M_o = x_k + \left\{h \times \frac{(f_k - f_{(k-1)})}{(2f_k - f_{k-1} - f_{k+1})}\right\}$$

$$= 40 + \left\{10 \times \frac{(28-12)}{(2\times 28 - 12 - 20)}\right\}$$

$$= 40 + \left\{10 \times \frac{16}{24}\right\} = \left(40 + \frac{20}{30}\right) = 40 + 6.67 = 46.67$$

Example 77: *Calculate the mode for the following frequency distribution:*

Class	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55
Frequency	25	34	50	42	38	14

Solution: The class 35 – 40 has maximum frequency. So, it is the modal class

∴ $x_k = 35, f_k = 50, f_{k-1} = 34, f_{k+1} = 42$ and $h = 5$

Mode, $$M_o = x_k + \left\{h \times \frac{(f_k - f_{k-1})}{(2f_k - f_{k-1} - f_{k+1})}\right\}$$

$$= 35 + \left\{5 \times \frac{(50 - 34)}{(2\times 50 - 34 - 42)}\right\} = 35 + \left\{5 \times \frac{16}{24}\right\}$$

$$= \left\{35 + \frac{10}{3}\right\} = (35 + 3.33) = 38.33$$

Problem Set

1. What is meant by central tendency ? Describe the various methods of measuring it and point out the usefulness of each method.

[B.C.A. (Bundelkhand) 2010, 2008, 2006; B.C.A. (Rohilkhand) 2009, 2004, 2000; B.C.A. (Avadh) 2008; B.C.A. (Garhwal) 2008, 2005]

2. Give a critical description of important measures of central tendency and their merits and demerits.

[B.C.A. (Meerut) 2011 2008, 2006; B.C.A. (Kanpur) 2009, 2005, 2003]

3. What is statistical average ? What are the desirable properties for an averages to possess ? Which of the averages, processes these properties ?

[B.C.A. (Garhwal) 2011, 2007; B.C.A. (Agra) 2009, 2005]

4. What do you mean by central tendency ? What are the desirable for an average ? Which average possess most of these properties ?

[B.C.A. (Kurukshetra) 2011; B.C.A. (Jodhpur) 2009, 2006; B.C.A. (Avadh) 2008, 2005; Kanpur 2007]

5. Write a short note on Geometrical mean and Harmonic mean.

[B.C.A. (Kanpur) 2008, 2005; B.C.A. (Agra) 2005]

6. From the following values locate the mode and median 33, 20, 35, 37, 33, 35, 25, 35, 34, 35.

[B.C.A. (Meerut) 2005]

7. Determine the mode in the following distribution:

Field in Circular form (In acre)	10	20	30	40	50	60	70
No. of Fields	7	12	17	29	31	5	3

8. Calculate mode by the method of grouping

Class-interval	Frequency
300–400	92
400–500	35
500–600	21
600–700	66
700–800	80
800–900	52
900–1000	20
1000–1100	14

[B.C.A. (Rohilkhand) 2006]

9. Find the median from the following distribution:

Size	18	20	21	19	23	22	24
Frequency	12	18	16	15	6	10	2

[B.C.A. (Kanpur) 2005]

10. Calculate the median and mode from the following series:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of Students	15	25	52	56	78	80	70

[B.C.A. (Meerut) 2007; B.C.A. (Rohilkhand) 2003]

11. Calculate Median and Mode from the following data:

Marks Less than	10	20	30	40	50	60	70
No. of Students	2	13	33	63	89	99	100

[B.C.A. (Meerut) 2004]

12. From the following table calculate the mode and first and third quartiles:

Income (₹)	100–200	100–300	100–400	100–500	100–600
No. of Persons	15	33	63	83	100

[B.C.A. (Rohilkhand) 2008; B.C.A. (Garhwal) 2004, 2006; B.C.A. (Meerut) 2005]

13. Five percent of the workers in a firm, employing of 2000 men, earn than ₹ 2.00 per hour, 480 earn from ₹ 2.00 to 22.24 per hour, 35% from ₹ 2.25 to ₹ 2.49 per hour, 370 earn from ₹ 2.50 to ₹ 2.75, 12% from ₹ 2.75 to ₹ 2.99 and the rest earn ₹ 3.00 or more per hour. What is the median wage ?

[B.C.A. (Udaipur) 2011, 2007]

14. In the following in complete distribution the value of median is 45, find out the missing frequencies:

Class-interval	10–20	20–30	30–40	40–50	50–60	60–70	70–80	Total
Frequencies	12	30	?	66	?	25	18	230

[B.C.A. (Rohilkhand) 2005, 2001]

15. Calculate arithmetic mean of following marks in Economics obtained by 10 students in a monthly test:

Roll Number	1	2	3	4	5	6	7	8	9	10
Marks	30	28	32	12	18	20	25	15	26	14

[B.C.A. (Avadh) 2009, 2005, 2003; B.C.A. (Kanpur) 2006, 2004]

16. Calculate the arithmetic mean and mode from the following table:

Mark (less than)	80	70	60	50	40	30	20	10
No. of Students	50	45	40	30	16	10	7	3

[B.C.A. (AMU) 2011, 2007; B.C.A. (Ajmer) 2009, 2005]

17.

(a) If the price of a commodity doubles in 4 years, what will be the average percentage ratio of increase per year ?

[B.C.A. (Meerut) 2006]

(b) The population of a country which was 30 crores in 1989 increased to 52 crores in 2007. Find the percentage compound rate of annual increase.

[B.C.A. (Udaipur) 2006]

18.

(i) A cyclist pedals from his house to his college at a speed of 10 km. p.h. and back from the college to his house at a speed of 15 km. p.h. What is his average speed ?

[B.C.A. (Agra) 2009, 2006]

(ii) A man travels from Lucknow to Meerut at a speed of 30 miles per hour and returns by the same route at 60 m.p.h. Find the average speed for the whole journey.

[B.C.A (Meerut) 2009, 2003, 2001; B.C.A. (Agra) 2005, 2003]

(iii) A train goes at a speed of 20 km. per hour for the first 16 kms, at a speed to 40 km. p.h. for 20 kms. It covers the last 10 kms at speed of 15 km.p.h. Find out its average speed.

[B.C.A (Udaipur) 2011, 2005; B.C.A. (AMU) 2009, 2007;

B.C.A. (Jabalpur) 2008, 2005]

19. Explain the difference between positional (vocational) average and mathematical average.

[B.C.A. (Kanpur) 2005, 2010]

20. The heights of 10 students are given below:

Students	A	B	C	D	E	F	G	H	I	J
Height (cms.)	156	154	169	161	163	167	165	181	158	166

Find the mean and median height.

21. The annual death-rates of a country for the last 14 years are given below:

10.6	10.9	11.1	10.6	10.9	11.6	10.4
10.6	11.1	10.7	10.7	10.6	10.9	11.3

Determine mode.

22. Calculate arithmetic mean, median and mode of the following frequency distribution of the size of 5 households:

Size	2	3	4	5	6	7
Number of Households	2	3	9	21	11	5

23. Find median Q_1, Q_3, D_6 and P_{30} of the following series:

Height (in inches)	58	59	60	61	62	63	64	65	66
No. of Persons	2	3	6	15	10	5	4	3	1

24. Find mean, median and mode from the following distribution:

Class	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Frequency	6	10	16	14	10	5	2

25. Find mean, median and mode from the following distribution:

Income (₹)	100–200	100–300	100–400	100–500	100–600
No. of Persons	15	33	63	83	100

[**Hint:** Convert the given distribution into a simple distribution]

26. Find median and quartiles from the following distribution:

Classes	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50
Frequency	7	10	13	26	35	22	11	5

[**Hint:** Obtain class boundaries before computation]

27. Determine mean, median and mode from the following table:

Marks (less than)	10	20	30	40	50	60	70	80
No. of Students	25	40	60	75	95	125	190	240

[**Hint:** Convert the given cumulative distribution into a simple one]

28. The mean height of 25 male workers in a factory is 61 inches and that of 35 female workers in the same factory is 58 inches. Find the average height of all the workers in the factory.

29. The mean annual salary of all the employees in a factory was ₹ 3000. The mean annual salaries paid to male and female employees are ₹ 3200 and ₹ 2200 respectively. Determine the percentage of male and female employed in the factory.

30. The sum of the deviations of a given observation from the A.M. is zero.

31. Define mean, median and mode with their merits and demerits. Why mean is consider as the best average ?

32. Find the mean, using direct method:

Class	0 – 10	10 – 20	20 – 30	40 – 50	50 – 60
Frequency	3	5	9	5	3

33.

Class	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500
Frequency	6	9	15	12	8

34. Find the missing frequencies f_1 and f_2 in the table given below, it is being given that the mean of the given frequency distribution is 50.

Class	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	Total
Frequency	17	f_1	32	f_2	19	120

35. Find the mean, using assumed-mean method:

(i)

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	12	18	27	20	17	6

(ii)

Class	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120
Frequency	20	35	52	44	38	31

(iii)

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	12	18	27	20	17	6

36. Calculate the mean for the following frequency distribution:

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	3	6	8	15	10	8

37. Calculate the median for the following data:

Age (in Years)	19 – 25	26 – 32	33 – 39	40 – 46	47 – 53	54 – 60
Frequency	35	96	68	102	35	4

38. Given below is the number of units of electricity consumed in a work in a certain locality.

Consumption (in Units)	65–85	85–105	105–125	125–145	145–165	165–185	185–205
No. of Consumers	4	5	13	20	14	7	4

39. The median value for the following frequency distribution is 35 and the sum of all the frequencies in 170. Using the formula for median, find the missing frequencies.

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	10	20	f_1	40	f_2	25	15

40. Compute the mode for the following distribution:

Class Interval	10 – 14	14 – 18	18 – 22	22 – 26	26 – 30	30 – 34	34 – 38	38 – 42
Frequency	8	6	11	20	25	22	10	4

41. Given below is the frequency distribution of the heights of the students in a school.

Height (in cm)	160 – 162	163 – 165	166 – 168	169 – 171	172 – 174
No. of Students	15	118	142	127	18

Find the average height of the maximum no. of students.

42. Mode of the following series is 36. Find the missing frequency in it.

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	8	10	?	16	12	6	7

43. Calculate the mode of the following data:

Monthly Salary (in ₹)	**No. of Employees**
0 – 5,000	90
5,000 – 10,000	150
10,000 – 15,000	100
15,000 – 20,000	80
20,000 – 25,000	70
25,000 – 30,000	10

Answers

6.	$Z = 35, M = 35$
7.	$Z = 41.46$ Acre
8.	$Z = 733.33$
9.	(i) $M = 8.5, Q_1 = 7.5, Q_3 = 9$ (ii) $M = 207.$
10.	$M = 45.128, Z = 51.667$
11.	$M = 45.128, Z = 37.14$
12.	$Z =$ ₹ $354.54, Q_1 = 255.56, Q_3 = 460$
13.	$M =$ ₹ 2.395
14.	40 and 39
15.	22 Marks
17.	(a) 19% (b) 3.1%
18.	(i) 12 km. p.h. (ii) 40 miles per hours (iii) 23.4 km.p.h.
20.	Mean = 164 cm. Median = 164 cm
21.	10.6
22.	Mean = Median = Mode = 5
23.	$Q_1 = 60$ inches, median = 61 inches, $Q_3 = 63$ inches, $P_{30} = 61$ inches $D_6 = 62$ inches
24.	Mean = 60.55, Median = 59.68, Mode = 57.50
25.	Mean = ₹ 347, Median = ₹ 356.67, Mode = ₹ 354.54
26.	Median = 31.71, $Q_1 = 25.93$, $Q_3 = 36.81$
27.	Mean = 49.58, Median = 58.33, Mode = 67.00
28.	59.25
29.	80% Male, 20% Female
32.	25
33.	264
34.	$f_1 = 28, f_2 = 24.$
35.	(i) 28 (ii) 62.55 (iii) 28
36.	35.33
37.	36
38.	136.5
39.	$f_1 = 35, f_2 = 25$
40.	28.5
41.	167.35 cm
42.	10
43.	₹ 7727.27

❑❑❑

Chapter

5

Measures of Dispersion

5.1 Dispersion

[B.C.A. (Meerut) 2010, 2009]

Dispersion is the measure of the variation of the items.

Or

A measure of variation or dispersion describes the degree of scatter shown by observations and it usually measured as an average deviation about control value or by an order statistics.

5.1.1 Objective of Measuring Variability or Dispersion

[Kashi Vidhyapeeth 2011; B.C.A. (Meerut) 2010]

Measures of variation or dispersion are computed to serve the following objectives:

(i) For judging the reliability of averages: Measures of dispersion help us to study the extent or degree to which the data are scattered around its average on central value.

(ii) Comparison of distributions.

(iii) Useful for controlling variability.

(iv) Useful in further statistical analysis.

5.1.2 Properties of a Good Measures of Dispersion

[B.C.A. (Meerut) 2010]

Since a measures of dispersion is the average of second order, the properties of a good measure of dispersion are the same as those for a good measure of central value as studied in the last chapter, *i.e.* these are:

(i) It should be simple to calculate and easy to understand.

(ii) It should be rigidly defined.

(iii) Its computation is based on all the observations.

(iv) It should be capable of further algebraic treatment.

(v) It should not be affected by extreme items.

(vi) It should be least affected by sampling fluctuations.

5.1.3 Two Types of Measures of Dispersion

(i) Absolute Measures of Dispersion

(ii) Relative Measures of Dispersion

(i) Absolute Measure of Dispersion: Absolute Measures of Dispersion *are expressed in the same unit in which the observations are given.* Thus, absolute measures of dispersion are useful for comparing variation in two or more distributions where units of measurement are same. *Such measures are not suitable for comparing the variability of the distributions expressed in different units of measurement.*

(ii) Relative Measures of Dispersion: Relative Measures of Dispersion *are expressed as ratio or percentage or the coefficient of the absolute measures of dispersion.* In this way, relative measures are pure unit less number and are generally called **coefficient of dispersion.** *Relative measures are useful for comparing variability in two or more distributions where units of measurement may be different.*

5.2 Important Measures of Dispersion

[B.C.A. (Meerut) 2010]

Following are some important measures of dispersion:

1. Range
2. Interquartile Range and Quartile Deviation
3. Mean Deviation or Average Deviation
4. Standard Deviation

1. **Range:** Range is the simplest measures of dispersion. *For a given set of observations, the* **range** *is the difference between the largest and the smallest observation.* Thus

Range; $R = L - S$ ...(1)

Here, L = the largest observation.

S = the smallest observation.

R = the Range.

In case of **grouped data**, *the range is defined as the difference between the upper limit of the highest class and the lower limit of the smallest class.*

Coefficient of Range

Range is an absolute measure of dispersion which is unsuitable for comparing variation in two or more distributions expressed in different units. To over come this difficulty, a relative measure of dispersion called the **coefficient of range** is calculated by the following formula:

$$\textbf{Coefficient of Range} = \frac{L - S}{L + S} \qquad ...(2)$$

Solved Examples

Example 1: *Marks of* 10 *students in Mathematics and Statistics are given below*:

Marks in Maths	25	40	30	35	21	45	23	33	10	29
Marks in Stat.	30	39	23	42	20	40	25	30	18	19

(*i*) *Compare the range of marks in the two subjects.*

(*ii*) *Compare the coefficients of range for both the subjects.*

Solution: The lowest and the highest marks in mathematics are 10 and 45 respectively. Also the lowest and highest marks in statistics are 18 and 42 respectively.

Thus, the range of marks in mathematics

$$R = L - S = 45 - 10 = 35$$

Range of marks in statistics

$$R = L - S = 42 - 18 = 24$$

$$\text{Coefficient of range (for Mathematics)} = \frac{L - S}{L + S} = \frac{45 - 10}{45 + 10} = \frac{35}{55} = 0.64$$

$$\text{Coefficient of range (for Statistics)} = \frac{L - S}{L + S} = \frac{42 - 18}{42 + 18} = \frac{24}{60} = 0.40$$

Thus, the range as well as the coefficient of range for marks in mathematics are higher than that of marks in statistics.

Example 2: *Find the range of the following distribution of families according to its size (No. of children).*

Size of the Family	0	1	2	3	4	5
No. of Families	8	12	15	25	4	6

Solution: In this example the variable (X) is the size of the family (No. of children). The minimum and the maximum size being 0 and 5 respectively. Thus, the range $R = L - S = 5 - 0 = 5$

Example 3: *Find the range and coefficient of range from the following data*:

Mid Value	5	10	15	20	25	30	35
Frequency	7	5	8	12	8	9	8

Solution: Here, the given mid-values may be used to define class boundaries or true limits of various classes. Since the uniform class-interval is 5, thus the limits for the first class will be $(5 - 2.5, 5 + 2.5)$, *i.e.* $(2.5 - 7.5)$, that of second will be $(7.5 - 12.5)$. The limits for the remaining class are shown in the table.

Thus the range

$$R = \text{upper limit of the highest class-lower limit of the lowest class.}$$

$$= 37.5 - 2.5 = \mathbf{35.0}$$

Computation of Class Limits

Mid-Value X	Class Limits	Frequency f
5	2.5–7.5	7
10	7.5–12.5	5
15	12.5–17.5	8
20	17.5–22.5	12
25	22.5–27.5	8
30	27.5–32.5	9
35	32.5–37.5	8

Also, **coefficient of range** $= \dfrac{L - S}{L + S}$

$$= \frac{37.5 - 2.5}{37.5 + 2.5}$$

$$= \frac{35.0}{40.0}$$

$$= \mathbf{0.875}$$

Example 4: *Find the range and coefficient of range from the following distribution*:

Classes	6–10	11–15	16–20	21–25	26–30	31–35	36–40
Frequency	7	8	15	35	18	7	5

Solution: First convert the given inclusive class limits into continuous classes (exclusive classes) as shown in the table:

Computing Continuous Class Limits

Inclusive Classes	Continuous Classes	Frequency
6–10	5.5–10.5	7
11–15	10.5–15.5	8
16–20	15.5–20.5	15
21–25	20.5–25.5	35
26–30	25.5–30.5	18
31–35	30.5–35.5	7
36–40	35.5–40.5	5

Here, $L = 40.5$ and $S = 5.5$

∴ $R = 40.5 - 5.5 = \mathbf{35}$

Coefficient of range $= \frac{L - S}{L + S}$

$$= \frac{40.5 - 5.5}{40.5 + 5.5}$$

$$= \frac{35}{46} = \mathbf{0.76}$$

Example 5: *The smallest value in a set of observations is 20 with a range of 25. Find the largest observation and the coefficient of range.*

Solution: Given that, $R = 25, S = 20$. To find the largest observation L, we consider

$$R = L - S \text{ or } 25 - L = 20$$

or $L = 25 + 20 = \mathbf{45}$

∴ **Coefficient of Range** $= \frac{L - S}{L + S} = \frac{45 - 20}{45 + 20} = \frac{25}{65} = \mathbf{0.37}$

Merits of Range

(i) It is simple to understand.

(ii) It is easy to calculate and provides a broad picture of the scatteredness in data quickly.

Demerits of Range

(i) Its composition is not based on all the observations.

(ii) It is affected by extreme items. A single extreme value, either the maximum or minimum can affect the value of range.

(iii) It is very much influenced by sampling fluctuations.

(iv) Range is a crude measure of dispersion. It does not tell us about the variation in the observations relative to the average.

(v) Range can not be calculated for open end distributions.

2. Interquartile Range and Quartile Deviation: The range is based on two extreme items and it does not take into account the variation within the range for this reason interquartile range is defined.

Interquartile Range includes the middle fifty percent of the Distribution : In other words, *interquartile range is the difference between the third quartile* (Q_3) *and the first quartile* (Q_1) *i.e.*

$$\text{Interquartile range} = Q_3 - Q_1 \qquad \text{...(3)}$$

In some case quartile deviation (Q.D.) or semi-interquartile range is defined as the average amount by which the two quartiles differ from the median therefore

Quartile deviation or Semi-interquartile range

$$= \frac{Q_3 - Q_1}{2} \qquad \text{...(4)}$$

The Coefficient of quartile

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad \text{...(5)}$$

Example 6: *From the following information of wages of* 15 *workers, find interquartile range, quartile deviation and co-efficient of Q.D.*

S. No.	1	2	3	4	5	6	7	8
Wages (₹)	520	550	440	580	450	620	470	680
S.No.	9	10	11	12	13	14	15	
Wages (₹)	400	490	420	480	440	480	500	

Solution: Arrange the given wages in the ascending order as under 400, 420, 440, 440, 450, 470, 480, 480, 490, 500, 520, 550, 580, 620, 680.

$$Q_1 \text{ is the size of } \left(\frac{N+1}{4}\right) = \frac{16}{4} = 4\text{th item}$$

$$\therefore \qquad Q_1 = ₹\ 440$$

$$Q_3 \text{ is the size of } \frac{3\,(N+1)}{4} = \frac{3\,(15+1)}{4} = 12\text{th item}$$

$$\therefore \qquad Q_3 = ₹\,550$$

$$\therefore \qquad \textbf{Interquartile-Range} = Q_3 - Q_1 = 550 - 440 = ₹\ \mathbf{110}$$

$$\textbf{Quartile deviation (Q.D.)} = \frac{Q_3 - Q_1}{2} = \frac{550 - 440}{2} = ₹\,\mathbf{55}$$

$$\textbf{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{550 - 440}{550 + 440} = \frac{110}{990} = \mathbf{0.11}$$

Example 7: *Scores are*

C.I.	*f*
45—49	2
40—44	3
35—39	2
30—34	6
25—29	8
20—24	8
15—19	5
10—14	7
5—9	9

Find the quartile deviation. **[B.C.A. (Meerut) 2004]**

Solution: Arranging given data in ascending order. This is a case of inclusive series. Converting it into exclusive series.

C.I.	Frequency	Commutative Frequency
4.5—9.5	9	9
9.5—14.5	5	14
14.5—19.5	7	21
19.5—24.5	8	29
24.5—29.5	8	37
29.5—34.5	6	43
34.5—39.5	2	45
39.5—44.5	3	48
44.5—49.5	2	50
	$N = 50$	

$$Q.\text{ No.} = \left(\frac{N}{4}\right)\text{th item}$$

$$= \frac{50}{4} = 12.5 \text{ th term which lies in class-interval } 9.5 - 14.5$$

$$Q_3 . \text{ No} = \left(\frac{3N}{4}\right)\text{ th item}=37.5\text{th lies in } 29.5 - 34.5$$

$$Q_1 = L_1 + \frac{L_2 - L_1}{f}\left(\frac{N}{4} - c\right)$$

$$Q_1 = 9.5 + \frac{14.5 - 9.5}{5}(12.5 - 9)$$

$$Q_1 = 9.5 + 3.5 = 13$$

$$Q_3 = 29.5 + \frac{34.5 - 29.5}{6}(37.5 - 37) = 29.917$$

$$\text{Quartile deviation} = \frac{(Q_3 - Q_1)}{2} = \frac{(29.917 - 13)}{2} = 8.4585$$

Example 8: *Calculate Q.D. and its co efficient from the following distribution:*

Weekly Income (₹)	58	59	60	61	62	63	64	65	66
No. of Workers	2	3	6	15	10	5	4	3	1

Solution:

Computing Q_1 and Q_3

Income X	**Workers f**	***c.f.***
58	2	2
59	3	5
60	6	11
61	15	26
62	10	36
63	5	41
64	4	45
65	3	48
66	1	49

Computation of Q_1 :

Q_1 is the size of $\frac{N+1}{4} = \frac{49+1}{4} = 12.5$th item.

$\therefore$ $\mathbf{Q_1 = ₹\,61}$

Computation of Q_3:

Q_3 is the size of $3\left(\frac{N+1}{4}\right) = 37.5$th item.

$$\therefore \quad \mathbf{Q_3 = ₹63}$$

$$\therefore \quad Q.D. = \frac{Q_3 - Q_1}{2} = \frac{63-61}{2}$$

$$= \mathbf{₹1.0}$$

$$\textbf{Coefficient of } \boldsymbol{Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{63-61}{63+61} = \frac{2}{124} = \mathbf{0.016}$$

Example 9: *The following is the age distribution of 799 workers:*

Age-Group	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60
No. of Workers	50	70	100	180	150	120	70	59

Find (i) Q.D. (ii) Coefficient of Q.D.

Solution: For determining Q.D. and its co efficient value, we first find the values of Q_1 and Q_3 from the given distribution.

Computation of Q_1 :

Q_1 is the value of $\frac{N}{4} = \frac{799}{4} = 199.75$th item. Thus, 30 – 35 is the *I* quartile class.

Computation Q_1 and Q_3

Classes	No. of Workers (*f*)	*c.f.*
20–25	50	50
25–30	70	120
30–35	100	220
35–40	180	400
40–45	150	550
45–50	120	670
50–55	70	740
55–60	59	799

Using the interpolation formula

$$Q_1 = l_1 + \frac{\frac{N}{4} - c}{f}(h)$$

$$= 30 + \frac{199.75 - 120}{100} \times 5$$

$$= 30 + \frac{398.75}{100}$$

$$= 30 + 3.99 = 33.99 \simeq \mathbf{34\ years.}$$

Computation of Q_3 :

Q_3 is the value of $\frac{3N}{4} = \frac{3 \times 799}{4} = 599.25$ th item which lies in $45 - 50$ class. Using the interpolation formula.

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c}{f} \times h$$

$$= 45 + \frac{599.25 - 550}{120} \times 5 = 45 + \frac{5 \times 49.25}{120}$$

$$= 45 + 2.05 = 47.05 \approx \mathbf{47\ years.}$$

$$\therefore \qquad \textbf{Q. D.} = \frac{Q_3 - Q_1}{2} = \frac{47 - 34}{2} = \frac{13}{2} = \mathbf{6.5\ years.}$$

and

$$\text{coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{47 - 34}{47 + 34} = \frac{13}{81} = \mathbf{0.16}$$

Merits of Q.D.

(i) It is simple to calculate.

(ii) It is easy to understand.

(iii) It is useful for measuring variations in open-end distributions.

(iv) It is also very useful when extreme items are likely to affect the analysis.

Demerits of Q.D.

(i) The calculation of Q.D. is not based on all observations.

(ii) It is not possible to give it further algebraic treatment.

(iii) It is very much affected by sampling fluctuations.

3. **Mean Deviation or Average Deviation:** Mean deviation *of a series is the arithmetic mean of the absolute deviations of various items from some central value, such as mean, median or mode.* Since some of the deviations about mean are positive and others negative, their algebraic sum about mean would be zero. That is why, is defining mean deviation, the average of the absolute [ignoring plus (+) and minus (–) signs] deviations from some central value is taken. In general, the central value can be any measure of central tendency, but usually the term mean deviation indicates

average of absolute deviations taken from mean. *Symbolically, the mean deviation about mean, median or mode can be expressed as follows:*

(a) Mean deviation from mean $\overline{X}$:

$$\text{M.D.}_{\overline{X}} = \frac{1}{N} \Sigma | X - \overline{X} | \qquad ...(6)$$

(b) Mean deviation from median $\mathbf{M_d}$:

$$\text{M.D.}_{M_d} = \frac{1}{N} \Sigma | X - M_d | \qquad ...(7)$$

(c) Mean deviation from mode $\mathbf{M_0}$:

$$\text{M.D.}_{M_0} = \frac{1}{N} \Sigma | X - M_0 | \qquad ...(8)$$

Here, N = the number of observations.

$\Sigma | X - \overline{X} |$ = Sum of absolute deviations taken from mean.

$\Sigma | X - M_d |$ = Sum of absolute deviations taken from median.

$\Sigma | X - M_0 |$ = Sum of absolute deviations taken from mode.

In case of grouped data or frequency distributions, the above formulae can be put as:

$$\text{M.D.}_{\overline{X}} = \frac{1}{N} \Sigma f | X - \overline{X} | \qquad ...(9)$$

$$\text{M.D.}_{M_d} = \frac{1}{N} \Sigma f | X - M_d | \qquad ...(10)$$

$$\text{M.D.}_{M_0} = \frac{1}{N} \Sigma f | X - M_0 | \qquad ...(11)$$

Coefficient of Mean Deviation

Mean deviation is an absolute measure of dispersion. The corresponding **relative measure called coefficient of mean deviation** *is obtained by dividing mean deviation by the average or central value used for calculating it.* Thus, symbolically,

$$\textbf{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Mean or Median or Mode}} \qquad ...(12)$$

In computing coefficient of M.D., it is important to note that the average used in calculating M.D. should be used in the denominator of formula (12). This relative measure of dispersion is useful for comparing two or more series which are expressed in different units or their averages are widely different.

Computation of Mean Deviation

1. Individual Series

Example 10: *Compute mean deviation and its coefficient from the following data relating to the marks obtained by a batch of* 11 *students in a class test*:

Marks	10	70	50	53	20	95	55	42	60	48	80

Solution:

Computation of M.D. (Individual Series)

X	Absolute Deviation from Mean $\lvert X - \overline{X} \rvert$
10	43
70	17
50	3
53	0
20	33
95	42
55	2
42	11
60	7
48	5
80	27
$\Sigma X = 583$	$\Sigma \lvert X - \overline{X} \rvert = 190$

Here, $$\overline{X} = \frac{\Sigma X}{N} = \frac{583}{11} = 53$$

∴ $$\text{M.D.} = \frac{1}{N} \Sigma \lvert X - \overline{X} \rvert$$

$$= \frac{190}{11} = \mathbf{17.27}$$

Also, on using formula (12), the coefficient of mean deviation is

$$= \frac{\text{M.D.}}{\text{Mean}} = \frac{17.27}{53} = \mathbf{0.326}$$

Steps

(i) Find Mean of the Series.

(ii) Obtain absolute deviations of observations from mean, *i.e.,* $| X - \overline{X} |$

(iii) Obtain the sum of absolute deviations, *i.e.* $\Sigma | X - \overline{X} |$

(iv) Use formula (6) to get M.D.

2. Discrete Series:

Example 11: *Calculate mean deviation from median for the following data. Also compute the coefficient of M.D.:*

Size	2	4	6	8	10	12	14	16
Frequency	2	2	4	5	3	2	1	1

Determination of Median

Size	***f***	***c.f.***
2	2	2
4	2	4
6	4	8
8	5	13
10	3	16
12	2	18
14	1	19
16	1	20

Solution: First we determine median of the given discrete distribution. Median is the value of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item $= \frac{20+1}{2} = 10.5$ th item.

∴ **Median** $= M_d = 8$

Using formula (10), the M.D. from M_d is

$$\text{M.D.} = \frac{1}{N} \Sigma f \, | X - M_d |$$

$$= \frac{56}{20} = 2.8$$

Computation of M.D. from Median

X	f	$\lvert X - M_d \rvert$	$f\lvert X - M_d \rvert$
2	2	6	12
4	2	4	8
6	4	2	8
8	5	0	0
10	3	2	6
12	2	4	8
14	1	6	6
16	1	8	8
	$N = 20$		$\Sigma f\lvert X - M_d\rvert = 56$

Using formula (12),

$$\text{The coefficient of M.D.} = \frac{M.D.}{M_d} = \frac{2.8}{8.0} = \mathbf{0.35}$$

3. Continuous Series

Example 12: *Compute the mean deviation (M.D.) from the following data.*

Classes	0–20	20–40	40–60	60–80	80–100	100–120
Frequency	5	50	84	32	10	6

Also find the coefficient of M.D.

Solution:

Computation of M.D. from Mean

Classes	X	f	$d' = \frac{X-50}{20}$	fd'	$\lvert X - \bar{X} \rvert$	$f\lvert X - \bar{X} \rvert$
0–20	10	5	– 2	– 10	41.07	205.35
20–40	30	50	– 1	– 50	21.07	1053.50
40–60	50	84	0	0	1.07	89.88
60–80	70	32	1	32	18.93	605.76
80–100	90	10	2	20	38.93	389.30
100–120	110	6	3	18	58.93	353.58
		$N = 187$		$\Sigma fd' = 10$		$\Sigma f\lvert X - \bar{X}\rvert = 2697.37$

$$\therefore \qquad \overline{X} = A + h\frac{\Sigma fd'}{N} = 50 + 20\frac{10}{187} = 50 + 1.07 = 51.07$$

$$\text{M.D.} = \frac{1}{N}\Sigma f\,|X - \overline{X}| = \frac{2697.37}{187} = \mathbf{14.42}$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Mean}} = \frac{14.42}{51.07} = \mathbf{0.28}$$

Merits of M.D.

(i) It is simple to understand.

(ii) It is easy to calculate.

(iii) Its computation is based on all the observations.

(iv) It is not affected by extreme observations.

Demerits of M.D.

(i) In the calculation of mean deviation, the algebraic signs of the deviations are ignored therefore its definition is non-algebraic. In view of this, mean deviation is not used in further statistical calculations.

(ii) If is not a well defined measure as any of the central value can be used in its computation. More so, the mean deviation calculated from different averages (mean, median or mode) will not be the same.

4. **Standard Deviation:** The concept of standard deviation was first introduced by Karl Pearson. It is one of the most popular and important measure of dispersion. It satisfies most of the properties of a good measure of dispersion. *The* **standard deviation** *is defined as the positive square root of the arithmetic mean of the squares of deviations of the observations from the arithmetic mean. It is also known as* '**root mean square deviation**' *and is generally denoted by the small Greek letter σ (called Sigma).* Symbolically for ungrouped data or individual series, the standard deviation is defined as

$$\textbf{S.D.} = \sigma = \sqrt{\left[\frac{1}{N}\Sigma\,(X - \overline{X})^2\right]} \qquad \text{...(13)}$$

In the case of a frequency distribution, the formula becomes

$$\sigma = \sqrt{\left[\frac{1}{N}\Sigma f\,(X - \overline{X})^2\right]} \qquad \text{...(14)}$$

The square of the standard deviation is known as variance. Therefore, variance for an individual series can be calculated by using the following formula:

$$\textbf{Variance} = \sigma^2 = \frac{1}{N}\Sigma\,(X - \overline{X})^2 \qquad \text{...(15)}$$

In case of grouped data or frequency distribution, the formula for calculating variance will be

$$\textbf{Variance} = \sigma^2 = \frac{1}{N}\Sigma f\,(X - \overline{X})^2 \qquad \text{...(16)}$$

Computation of Standard Deviation (S.D.) in Individual Series or Ungrouped Data

For ungrouped data, the following two methods are used for computing standard deviation:

(i) Direct Method

(ii) Short-cut Method

(i) Direct Method: In this procedure, the following formula is used for calculating S.D. (σ):

$$\sigma = \sqrt{\left[\frac{1}{N}\Sigma\,(X - \overline{X})^2\right]} \qquad \text{...(17)}$$

or

$$\sigma = \sqrt{\left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right]} = \sqrt{\left[\frac{\Sigma X^2}{N} - \overline{X}^2\right]} \qquad \text{... (18)}$$

Theorem: Show that for only discrete distribution their S.D. is not less than the mean deviation from the mean.

[B.C.A. (Meerut) 2004, 2002, 2001]

Proof: With usual notation we have

$$\sqrt{\frac{1}{N}\Sigma\, f\,(x - M)^2} \geq \left\{\frac{1}{N}\Sigma\, f\,|\,x - M\,|\right\}, M = \text{mean}$$

$$\Rightarrow \quad \frac{1}{N}\Sigma f\,(x - M)^2 \geq \left\{\frac{1}{N}\Sigma\, f\,|\,x - M\,|\right\}^2$$

$$\Rightarrow \quad N\Sigma\, f\,\xi_i^2 \geq (\Sigma f\xi_i)^2 \text{ where } \xi_i = |\,x - M\,|$$

$$(f_1 + f_2 + f_3 + \ldots + f_n)\,(f_1\,\xi_1^2 + f_2\,\xi_2^2 + \ldots + f_n\,\xi_n^2)$$

$$\geq (f_1\xi_1 + f_2\xi_2 + \ldots + f_n\xi_n)^2$$

$$\Rightarrow \quad f_1 f_2(\xi_1^2 + \xi_2^2 + \ldots \geq 2\, f_1 f_2\, \xi_1\xi_2$$

$$\Rightarrow \quad f_1 f_2(\xi_1 - \xi_2)^2 \ldots \geq 0$$

Being the sum of perfect squares. L.H.S. is never negative. Hence, the proposition.

Example 13: *Calculate S.D. from the following set of observations:*

$$X: 10, 11, 17, 25, 7, 13, 21, 10, 12, 14$$

Solution:

Computing S.D.

X	$(X-\overline{X})$	$(X-\overline{X})^2$
10	– 4	16
11	– 3	9
17	3	9
25	11	121
7	– 7	49
13	– 1	1
21	7	49
10	– 4	16
12	– 2	4
14	0	0
$\Sigma X = 140$		$\Sigma(X-\overline{X})^2 = 274$

Here, $$\overline{X} = \frac{\Sigma X}{N} = \frac{140}{10} = 14$$

Now, $$\text{S.D.} = \sigma = \sqrt{\left[\frac{1}{N}\Sigma(X-\overline{X})^2\right]}$$

$$= \sqrt{\left[\frac{1}{10} \times 274\right]}$$

$$= \sqrt{(27.4)} = \mathbf{5.23}$$

(ii) Short-Cut Method: In most of the cases, the arithmetic mean of the given distribution happens to be a fractional value and then the process of taking deviation and squaring them becomes quite tedious and time consuming in the computation of S.D. To overcome this difficulty, short-cut method of computation is used which involves deviations from assumed mean. The short-cut formula for calculating S.D. is:

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2\right]} \qquad \text{...(19)}$$

Here, d = deviation from assumed mean, say A; *i.e.* $d = (X - A)$

Σd = the sum of deviations

Σd^2 = the sum of squares of deviations

N = the number of observations.

Example 14: *Find S.D. from the data in example 1*

Solution:

Computation of S.D.

X	$d = X - A$	d^2
10	– 2	4
11	– 1	1
17	+ 5	25
25	+ 13	169
7	– 5	25
13	+ 1	1
21	+ 9	81
10	– 2	4
12	0	0
14	2	4
$N = 10$	$\Sigma d = 20$	$\Sigma d^2 = 314$

Let, $A = 12$

Using formula (19), *i.e.*

$$\sigma = \sqrt{\left[\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2\right]}$$

$$= \sqrt{\left[\frac{314}{10} - \left(\frac{20}{10}\right)^2\right]}$$

$$= \sqrt{[31.4 - 4]} = \sqrt{[27.4]}$$

$$= \mathbf{5.23}$$

Computation of S.D. in Grouped Data

Any of the following three procedures may be applied to find S.D. for grouped data :

1. Direct Method
2. Short-Cut Method
3. Step-Deviation Method

1. **Direct Method:** In direct method of computing S.D., the following formula is used:

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{1}{N}\Sigma f\,(X-\overline{X})^2\right]} \qquad ...(20)$$

where symbols have their usual meaning.

Example 15: *Use direct method to calculate S.D. of the following discrete frequency distribution:*

Size (X)	4	5	6	7	8	9	10
Frequency	6	12	15	28	20	14	5

Solution:

Computation of S.D. (Direct-Method)

Size X	Frequency f	fX	$(X-\overline{X})$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
4	6	24	– 3.06	9.3636	56.1816
5	12	60	– 2.06	4.2436	50.9232
6	15	90	– 1.06	1.1236	16.8540
7	28	196	– 0.06	0.0036	0.1008
8	20	160	+ 0.94	0.8836	17.6720
9	14	126	+ 1.94	3.7636	52.6904
10	5	50	+2.94	8.6436	43.2180
	$N = 100$	$\Sigma fx = 706$			$\Sigma f(X-\overline{X})^2 = 237.6400$

Here, we first calculate:

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{706}{100} = 7.06$$

Using formula (20),

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{1}{N}\Sigma f\,(X-\overline{X})^2\right]}$$

$$= \sqrt{\left(\frac{237.64}{100}\right)} = \sqrt{(2.3764)} = \mathbf{1.54}$$

2. **Short-Cut Method:** In the direct method of computing S.D., we face a difficult situation when the arithmetic mean is not a whole number. To avoid such difficulties, we consider the deviations of observations from a suitably chosen assumed mean, say *A*. In short-cut method of computing the S.D. we make use of the following formula:

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2\right]} \qquad ...(21)$$

where *d* is deviation from assumed mean. The short-cut method of calculating S.D. involves the following steps.

Example 16: *Use short-cut method to find the S.D. of the data in example 3.*

Computation of S.D. (Short-Cut Method)

X	f	$d = (X - 7)$	fd	d^2	fd^2
4	6	– 3	– 18	9	54
5	12	– 2	– 24	4	48
6	15	– 1	– 15	1	15
7	28	0	0	0	0
8	20	1	+ 20	1	20
9	14	2	+ 28	4	56
10	5	3	+15	9	45
	$N = 100$		$\Sigma fd = 6$		$\Sigma fd^2 = 238$

Solution:

Using formula (21),

$$\sigma = \sqrt{\left[\frac{\Sigma f\,d^2}{N} - \left(\frac{\Sigma f\,d}{N}\right)^2\right]} = \sqrt{\left[\frac{238}{100} - \left(\frac{6}{100}\right)^2\right]}$$

$$= \sqrt{[2.38 - 0.0036]} = \sqrt{2.3764} = \mathbf{1.54}$$

3. **Step-Deviation Method:** In short-cut method of calculation our main aim is to simplify the deviations so that computations become easier. In grouped data, specially in continuous frequency distributions, we observe that the calculations can be further simplified if deviations (d) are divided by a common factor, say *h*, which is usually the size of the class-interval to get the step-deviations as *d′*. However, such a division is then reflected in the formula for computing S.D. which now becomes:

$$\sigma = h\sqrt{\left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right]} \qquad ...(22)$$

Steps involved in computing S.D. by using step-deviation method will be clear from the following example:

Example 17: *Use short-cut and step-deviation method for calculating S.D. of the following distribution.*

Age Groups (years)	25–30	30–35	35–40	40–45	45–50	50–55
No. of Workers	10	12	25	40	10	3

Solution: **Using Short-Cut Method.**

Computation of S.D. (Short-Cut Method)

Age	Frequency f	Mid Value X	$d = (X - 42.5)$	fd	d^2	fd^2
25–30	10	27.5	– 15	– 150	225	2250
30–35	12	32.5	– 10	– 120	100	1200
35–40	25	37.5	– 5	– 125	25	625
40–45	40	42.5	0	0	0	0
45–50	10	47.5	5	50	25	250
50–55	3	52.5	10	30	100	300
	$N = 100$			$\Sigma fd = -315$		$\Sigma fd^2 = 4625$

The assumed mean is taken as $A = 42.5$. Using formula (22).

$$\sigma = \sqrt{\left[\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2\right]} = \sqrt{\left[\frac{4625}{100} - \left(\frac{-315}{100}\right)^2\right]}$$

$$= \sqrt{[46.25 - 9.9225]} = \sqrt{[36.3275]} = \textbf{6.03 years.}$$

Using Step-Deviation Method

Computation of S.D. (Step-Deviation Method)

Age	Frequency f	Mid-Value X	$d' = \frac{X - 42.5}{5}$	fd'	fd'^2
25–30	10	27.5	– 3	– 30	90
30–35	12	32.5	– 2	– 24	48
35–40	25	37.5	– 1	– 25	25
40–45	40	42.5	0	0	0
45–50	10	47.5	1	10	10
50–55	3	52.5	2	6	12
	$N = 100$			$\Sigma fd' = -63$	$\Sigma fd'^2 = 185$

Here, $A = 42.5$ and $h = 5$. Using formula (22), *i.e.*

$$\sigma = h\sqrt{\left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right]}$$

Putting the values from the table,

$$\sigma = 5\sqrt{\left[\frac{185}{100} - \left(\frac{63}{100}\right)^2\right]} = 5\sqrt{[(1.85 - 0.3969)]}$$

$$= 5\sqrt{(1.4531)} = 5 \times 1.2054 = \mathbf{6.03\ years.}$$

which is the same as that obtained by short-cut method.

Example 18: *From the following information, find the S.D.*

$$N = 10, \Sigma X = 60, \Sigma X^2 = 1000$$

[B.C.A. (Rohilkhand) 2006]

Solution: Using the formula (18), *i.e.*

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right]} = \sqrt{\left[\frac{1000}{10} - \left(\frac{60}{10}\right)^2\right]}$$

$$= \sqrt{[100 - 36]} = \mathbf{8.00}$$

Example 19: *Use the following information to find the variance.*

$$N = 20, \ \overline{X} = 50, \ \Sigma X^2 = 84000$$

[B.C.A. (Agra) 2004]

Solution: Using formula (18), *i.e.*

$$\text{S.D.} = \sigma = \sqrt{\left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right]} = \sqrt{\left[\frac{\Sigma X^2}{N} - \overline{X}^2\right]}$$

Squaring on both sides:

$$\text{Variance} = \sigma^2 = \left[\frac{\Sigma X^2}{N} - \overline{X}^2\right]$$

Putting the given values, one gets

$$\sigma^2 = \frac{84000}{20} - (50)^2 = 4200 - 2500 = \mathbf{1700.00}$$

Example 20: *Find the value of S.D. and the variance from the following information*:

$$\Sigma\,(X - 7) = 8, \Sigma\,(X - 7)^2 = 535, \ N = 15$$

Solution: For computing the S.D. or the variance, we need the values of N, ΣX and ΣX^2 which can be gathered from the given information.

Since $\Sigma\,(X - 7) = 7$

or $\Sigma X - \Sigma 7 = 7$

or $\Sigma X - 15 \times 7 = 7$ [Using $\Sigma c = Nc$]

∴ $\Sigma X = 105 + 7 = 112$ ...(1)

Again $\Sigma (X - 7)^2 = 535$

or $\Sigma (X^2 - 14X + 49) = 535$ [Using $\Sigma\, cX = c\Sigma X$]

or $\Sigma X^2 - 14\Sigma X + \Sigma 49 = 535$

or $\Sigma X^2 - 14 \times 112 + 15 \times 49 = 535$

$\Sigma X^2 = 535 + 14 \times 112 - 15 \times 49 = 1368.$ ...(2)

Using (1) and (2) the variance

$$\sigma^2 = \left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right] = \left[\frac{1368}{15} - \left(\frac{112}{15}\right)^2\right]$$

$$= 91.20 - 55.75 = \mathbf{35.45}$$

∴ $$\text{S.D.} = \sqrt{\sigma^2} = \sqrt{(35.45)} = \mathbf{5.95}$$

Correcting Incorrect Values of Mean and Standard Deviation

Important Remarks:

(i) ΣX can be obtained if N and $\overline{X}$ are given.

(ii) *ΣX^2 can be computed if N, $\overline{X}$ and σ are mentioned.*

(iii) *It was observed in the last chapter that for correcting the incorrect mean value, we need to find the corrected ΣX and ΣfX. For this we subtract the wrong items from the incorrect ΣX or ΣfX and add to it the correct items.*

(iv) *For correcting the incorrect value of S.D. or variance, we need the corrected values of ΣX and ΣX^2 both. Corrected ΣX^2 (or ΣfX^2) is obtained by subtracting the squares of the wrong items from the incorrect ΣX^2 (or ΣfX^2) and adding the squares of the correct items.*

The following example will clarify these points.

Example 21: *The mean of* 100 *items is* 50 *and their S.D. is* 4. *Find the sum and the sum of square of the items.*

Solution:

(i) Computing sum of the items, *i.e.* ΣX :

Given that $N = 100$ and $\overline{X} = 50$

∴ The sum of the items = $\Sigma X = N\overline{X} = 100 \times 50 = \mathbf{5000}$

(ii) Computing sum of squares of the items, ***i.e.***, ΣX^2 :

$$\text{Given that } N = 100, \overline{X} = 50, \Sigma X = 5000 \text{ and } \sigma = 4.$$

$$\text{Also knowing that S.D.} = \sigma = \sqrt{\left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right]}$$

Putting the given values, we have

$$4 = \sqrt{\left[\frac{\Sigma X^2}{100} - \left(\frac{5000}{100}\right)^2\right]}$$

Squaring on both sides,

$$16 = \left[\frac{\Sigma X^2}{100} - 2500\right]$$

or $$2500 + 16 = \frac{\Sigma X^2}{100}$$

or $$\Sigma X^2 = 100 \times [2516] = \mathbf{251600}$$

Example 22: *The sum of squares of 20 observations was worked out as* 5100. *But while calculating it, an observation* 31 *was wrongly considered as* 13. *Find the corrected sum of squares.*

Solution: Given that incorrect sum of squares, *i.e.* $\Sigma X^2 = 5100$. Now, the corrected ΣX^2 = incorrect ΣX^2 – (Square of the wrong item) + (Square of the correct item) $= 5100 - (13)^2 + (31)^2 = 5100 - 169 + 961 = \mathbf{5892}$ **.**

Example 23: *The sum of squares of* 100 *observations was calculated as* 7961. *Later, it was found that two values, 53 and 42 were wrongly read as 35 and 24 at the time of calculation. Find the corrected sum of squares.*

Solution: Given that incorrect $\Sigma X^2 = 7961$.

$\therefore$ Corrected ΣX^2 = incorrect ΣX^2 – (Squares of wrong observations)

+ (Squares of correct observations)

$$\therefore \quad \text{Corrected } \Sigma X^2 = 7961 - (35^2 + 24^2) + (53^2 + 42^2)$$

$$= 7961 - (1225 + 576) + (2809 + 1764) = \mathbf{10733}$$

Example 24: *The sum of squares of* 50 *observations is* 4122. *An observation* 39 *was wrongly included in the series. Find the sum of squares of the remaining* 49 *observations.*

Solution: Incorrect $\Sigma X^2 = 4122$.

$\therefore$ Corrected ΣX^2 = incorrect ΣX^2 – (Squares of the wrongly taken observation)

$$= 4122 - (39^2) = 4122 - 1521 = \mathbf{2601}$$

Example 25: *The arithmetic mean and the S.D. of a series of* 20 *items were calculated as* 20 *cm. and* 5 *cm. respectively. But while calculating them, an item* 13 *was measured as* 30. *Find the correct arithmetic mean and standard deviation.* **[B.C.A. (Meerut) 2008, 2006]**

Solution: Given that $N = 20$, $\overline{X} = 20$, $\sigma = 5$, incorrect observation=30, correct observation=13.

Thus, in corrected $\Sigma X = N\overline{X} = 20 \times 20 = 400$

$\therefore$ Corrected ΣX = incorrected ΣX – (incorrect observation) + (correct observation)

$$= 400 - 20 + 31 = \mathbf{411}$$

$$\therefore \quad \text{Corrected Mean} = \frac{\text{Corrected } \Sigma X}{N} = \frac{411}{20} = \mathbf{20.55} \qquad \ldots(1)$$

For finding the corrected S.D., we first find the incorrect ΣX^2 from the given $\overline{X}$ and σ values by using the formula

$$\sigma^2 = \left[\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2\right] = \frac{\Sigma X^2}{N} - \overline{X}^2$$

or $$\Sigma X^2 = N(\sigma^2 + \overline{X}^2)$$

$$\therefore \quad \text{incorrect } \Sigma X^2 = N(\sigma^2 + \overline{X}^2) = 20\,(25 + 400) = 8500$$

Thus, corrected ΣX^2 = incorrect ΣX^2 – (Squares of incorrect values) + (Squares of correct values)

$$= 8500 - 30^2 + 13^2 = 8500 - 900 + 169 = \mathbf{7769} \ \ldots(2)$$

Using (1) and (2), the corrected S.D. becomes

$$\text{Corrected S.D.} = \sqrt{\left[\frac{\text{corrected } \Sigma X^2}{N} - \left(\frac{\text{corrected } \Sigma X}{N}\right)^2\right]}$$

$$= \sqrt{\left[\frac{7769}{20} - \left(\frac{411}{20}\right)^2\right]} = \sqrt{388.45 - 422.30}$$

$$= \sqrt{(21.73)} = \mathbf{4.66}$$

Example 26: *The mean and S.D. of* 20 *items are found to be* 10 *and* 2 *respectively. At the time of checking, it was observed that one item* 8 *was incorrect. Find the mean and the S.D. if* (i) *the wrong item is omitted* (ii) *it is replaced by* 12.

Solution: Given that $N = 20$, $\overline{X} = 10$, $\sigma = 2$ and wrong item = 8.

Thus, $$\text{incorrect } \Sigma X = N\overline{X} = 20 \times 10 = 200$$

$$\text{incorrect } \Sigma X^2 = N(\sigma^2 + \overline{X}^2) = 20\,(4 + 100) = 2080$$

Case (i) when the wrong item is omitted

$$\text{Corrected } N = \text{incorrect } N - 1 = 20 - 1 = 19$$

$$\text{Corrected } \Sigma X = \text{incorrect } \Sigma X - \text{wrong item} = 200 - 8 = 192$$

$$\text{Corrected } \Sigma X^2 = \text{incorrect } \Sigma X^2 - \text{(Square of the wrong item)}$$

$$= 2080 - 8^2 = 2080 - 64 = 2016$$

$$\therefore \qquad \textbf{Corrected Mean} = \frac{\text{corrected } \Sigma X}{\text{corrected } N} = \frac{192}{19} = \mathbf{10.11}$$

$$\text{Corrected S.D.} = \sqrt{\left[\frac{\text{corrected } \Sigma \text{X}^2}{\text{corrected N}} - \left(\frac{\text{corrected } \Sigma\text{X}}{\text{corrected N}}\right)^2\right]}$$

$$= \sqrt{\left[\frac{2016}{19} - \left(\frac{192}{19}\right)^2\right]} = \sqrt{[106.11 - 102.12]}$$

$$= \sqrt{(3.99)} = \mathbf{1.99} \approx \mathbf{2.00}$$

Case (ii) when the wrong item is replaced by 12.

$$\text{Corrected } \Sigma X = \text{incorrect } \Sigma X - \text{(Wrong item)} + \text{(Correct item)}$$

$$= 200 - 8 + 12 = 204$$

$$\therefore \qquad \text{Corrected } \overline{X} = \frac{\text{corrected } \Sigma X}{N} = \frac{204}{20} = \mathbf{10.2}$$

$$\text{Corrected } \Sigma X^2 = \text{incorrected } \Sigma X^2 - \text{(Square of wrong item)}$$

$$+ \text{(Square of correct item)}$$

$$= 2080 - 64 + 144 = 2160$$

$$\text{Thus,} \qquad \textbf{corrected S.D.} = \sqrt{\left[\frac{\text{corrected } \Sigma X^2}{N} - \left(\frac{\text{corrected } \Sigma X}{N}\right)^2\right]}$$

$$= \sqrt{\left[\frac{2160}{20} - \left(\frac{204}{20}\right)^2\right]} = \sqrt{[108 - 104.04]}$$

$$= \sqrt{[3.96]} = \mathbf{1.99}$$

Algebraic Properties of Standard Deviation

We observed that if the means and the number of items in different groups are given, the combined mean of all the groups can be computed. Similarly, combined standard deviation can be calculated if the standard deviations, means and number of items in different groups are given. The formula used for computing combined standard deviation is as under:

1. **Combined S.D.:**

(i) For two Related Groups:

$$\sigma_{12} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]} \quad \text{...(23)}$$

where, σ_{12} = Combined S.D. of two groups

σ_1 = Standard deviation of the first group

σ_2 = Standard deviation of the second group

N_1 = No. of observations in the first group.

N_2 = No. of observations in the second group.

$D_1 = (\overline{X}_1 - \overline{X}_{12})$, $\overline{X}_{12}$ = combined mean of the two groups.

$D_2 = (\overline{X}_2 - \overline{X}_{12})$.

$\overline{X}_1$ = mean of the first group.

$\overline{X}_2$ = mean of the second group.

(ii) For three Related Groups:

$$\sigma_{123} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2) + N_3(\sigma_3^2 + D_3^2)}{(N_1 + N_2 + N_3)}\right]} \quad \text{...(24)}$$

Here, σ_{123} = Combined standard deviation of three groups.

$D_1 = (\overline{X}_1 - \overline{X}_{123}), D_2 = (\overline{X}_2 - \overline{X}_{123})$

$D_3 = (\overline{X}_3 - \overline{X}_{123})$, $\overline{X}_{123}$ being the combined mean of the three groups.

Other symbols have their usual meanings. The formula for combined S.D. can easily be extended to any number of groups.

2. **Standard Deviation and Normal Curve**: For a normal distribution or a symmetrical distribution, the number of observations lying within specific ranges can be measured with a greater precision. In such distributions the percentage of observations lying in some specific ranges are as under:

(i) $\overline{X} \pm \sigma$, also called 1-sigma limits, covers 68.27% of the total items.

(ii) $\overline{X} \pm 2\sigma$, also called 2-sigma limits, covers 95.45% of the total items.

(iii) $\overline{X} \pm 3\sigma$, also called 3-sigma limits, covers 99.73% of the total items.

3. **Empirical Relationship between Measures of Dispersion:** For a normal distribution or for a moderately a symmetrical distribution, following empirical relationships exist between various measure of dispersion.

(i) Quartile Deviation (Q.D.) $= \frac{2}{3}$ S.D.

(ii) Mean Deviation (M.D.) $= \left(\frac{4}{5}\right)$ S.D.

(iii) Quartile Deviation (Q.D.) $= \left(\frac{5}{6}\right)$ M.D.

(iv) 6 S.D. = 9 Q.D. = 7.5 M.D.

Example 27: *A sample of 35 values has mean 80 and S.D. 4. A second sample of 65 values from the same population has mean 70 and S.D. 3. Find the mean and standard deviation of the combined sample of 100 values.*

Solution: Given that $N_1 = 35, \overline{X}_1 = 80, \sigma_1 = 4;$

$$N_2 = 65, \overline{X}_2 = 70, \sigma_2 = 3$$

Combined Mean:

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{35 \times 80 + 65 \times 70}{100} = \frac{2800 + 4550}{100} = \mathbf{73.50}$$

Combined S.D. :

$$D_1 = (\overline{X}_1 - \overline{X}_{12}) = (80 - 73.5) = +6.5 \quad \therefore \; D_1^2 = 42.25$$

$$D_2 = (\overline{X}_2 - \overline{X}_{12}) = (70 - 73.5) = -3.5 \quad \therefore \; D_2^2 = 12.25$$

$$\therefore \quad \sigma_{12} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]}$$

$$= \sqrt{\left[\frac{35(16 + 42.25) + 65(9 + 12.25)}{35 + 65}\right]} = \sqrt{\left[\frac{2038.75 + 1381.25}{100}\right]}$$

$$= \sqrt{[34.20]} = \mathbf{5.85}$$

Example 28: *Find the mean and standard deviation of the two groups taken together:*

Group	*Number*	*Mean*	*S.D.*
A	113	160	22
B	120	150	20

Solution: Given that

$$N_1 = 113, \overline{X}_1 = 160, \sigma_1 = 22; N_2 = 120, \overline{X}_2 = 150, \sigma_2 = 20.$$

Combined Mean :

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{113 \times 160 + 120 \times 150}{113 + 120} = \frac{18080 + 18000}{233}$$

$$= \frac{36080}{233} = \mathbf{154.85}$$

Combined S.D.:

$$D_1 = (\overline{X}_1 - \overline{X}_{12}) = (160 - 154.85) = 5.15 \quad \therefore\ D_1^2 = 26.5225$$

$$D_2 = (\overline{X}_2 - \overline{X}_{12}) = (150 - 154.85) = -4.85 \quad \therefore\ D_2^2 = 23.5225$$

$$\therefore \qquad \sigma_{12} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]}$$

$$= \sqrt{\left[\frac{113\,(484 + 26.5225) + 120\,(400 + 23.5225)}{113 + 120}\right]}$$

$$= \sqrt{\left[\frac{57689.04 + 50822.70}{233}\right]} = \sqrt{[465.72]} = \mathbf{21.58}$$

Example 29: *The following data were recorded in two sample studies:*

	Sample I	***Sample II***
No. of Observations	150	100
Mean	40	55
Variance	26	35

Calculate the combined mean and S.D. of the observations in the two sample together.

[B.C.A. (Meerut) 2002]

Solution: The following data are given for the two samples:

Sample-1	$N_1 = 150$	$\overline{X}_1 = 40$	$\sigma_1^2 = 26$
Sample-2	$N_2 = 100$	$\overline{X}_2 = 55$	$\sigma_2^2 = 35$

Combined Mean :

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{150 \times 40 + 100 \times 55}{150 + 100} = \frac{11500}{250} = \mathbf{46.00}$$

For combined S.D., we first compute—

$$D_1 = (\overline{X}_1 - \overline{X}_{12}) = 40 - 46 = -6.0 \quad \therefore\ D_1^2 = 36.0$$

$$D_2 = (\overline{X}_2 - \overline{X}_{12}) = 55 - 46 = 9.0 \quad \therefore\ D_2^2 = 81.0$$

Putting all these values in the formula for combined S.D., *i.e.*

$$\sigma_{12} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]}$$

$$= \sqrt{\left[\frac{150\,(26 + 36) + 100\,(35 + 81)}{150 + 100}\right]} = \sqrt{\left[\frac{9300 + 11600}{250}\right]} = \mathbf{9.14}$$

Example 30: *The mean of the two samples of sizes* 50 *and* 100 *respectively is* 54.1 *and* 50.3 *and standard deviations are* 8 *and* 7. *Obtain the mean and standard deviation of combined sample of size* 150.

[B.C.A. (Meerut) 2001]

Solution: Given

$$N_1 = 50,\ N_2 = 100$$
$$\overline{X}_1 = 54.1,\ \overline{X}_2 = 50.3$$
$$\sigma_1 = 8,\qquad \sigma_2 = 7$$

Combined Mean $(\overline{X}_{12}) = \dfrac{N_1\ \overline{X}_1 + N_2\ \overline{X}_2}{N_1 + N_2} = \dfrac{50 \times 54.1 + 100 \times 50.3}{150} = 51.57$

Combined Standard Deviation

$$\sigma_{12} = \frac{\sqrt{N_1\ \sigma_1^2 + N_2\ \sigma_2^2 + N_1\ (\overline{X}_1 - \overline{X}_{12})^2 + N_2\ (\overline{X}_2 - \overline{X}_{12})^2}}{N_1 + N_2}$$

$$= \frac{\sqrt{50 \times 64 + 100 \times 49 + 50 \times 6.4 + 100 \times 1.6129}}{150}$$

$$= \frac{\sqrt{3200 + 4900 + 320 + 161.20}}{150}$$

$$\sigma_{12} = 7.564$$

Example 31: *The following data were recorded for three distributions:*

Distributions	*N*	$\overline{X}$	*S.D.*
A	20	60	8
B	50	50	12
C	60	40	20

Obtain combined mean and S.D. of all the three distributions together.

Solution: For the three distributions, the data can be put as:

Distribution A	$N_1 = 20,$	$\overline{X}_1 = 60,$	$\sigma_1 = 8$
Distribution B	$N_2 = 50,$	$\overline{X}_2 = 50,$	$\sigma_2 = 12$
Distribution C	$N_3 = 60,$	$\overline{X}_3 = 40,$	$\sigma_3 = 20$

Combined Mean:

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3} = \frac{20 \times 60 + 50 \times 50 + 60 \times 40}{20 + 50 + 60}$$

$$= \frac{6100}{130} = \mathbf{46.92}$$

For combined S.D., we first calculate:

$$D_1 = (\overline{X}_1 - \overline{X}_{123}) = 60 - 46.92 = 13.08 \quad \therefore\ D_1^2 = 171.09$$

$$D_2 = (\overline{X}_2 - \overline{X}_{123}) = 50 - 46.92 = 3.08 \quad \therefore\ D_2^2 = 9.49$$

$$D_3 = (\overline{X}_3 - \overline{X}_{123}) = 40 - 46.92 = -6.92 \quad \therefore\ D_3^2 = 47.89$$

Also given $\sigma_1 = 8$, $\sigma_2 = 12$ and $\sigma_3 = 20$. Thus, the combined S.D. will be

$$\sigma_{123} = \sqrt{\left[\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2) + N_3(\sigma_3^2 + D_3^2)}{N_1 + N_2 + N_3}\right]}$$

$$= \sqrt{\left[\frac{20(64 + 171.09) + 50(144 + 9.49) + 60(400 + 47.89)}{20 + 50 + 60}\right]}$$

$$= \sqrt{\left[\frac{4701.80 + 7674.50 + 26873.40}{130}\right]} = \sqrt{\left[\frac{39249.70}{130}\right]}$$

$$= \sqrt{[301.92]} = \mathbf{17.38}$$

Merits of S.D.

(i) It is rigidly defined.

(ii) Its computation is based on all the observations.

(iii) It is amenable to further algebraic treatment which makes it the most important and widely used measure of dispersion. For example, the S.D. is used in computing skewness, correlation etc. It is an important statistical measure in sampling theory.

(iv) Among all the measures of dispersion, it is least affected by sampling fluctuations.

(v) S.D. enables us to determine the reliability of means of two or more series having equal means. In such a situation, a series having minimum S.D. will have the most representative mean. In other words, a smaller value of S.D. reflect greater compactness and smaller variability among items.

Demerits of S.D.

(i) S.D. is comparatively difficult to calculate.

(ii) It gives greater weight to extreme observations.

(iii) It is an absolute measure of dispersion and cannot be used for comparing variability of two or more distributions expressed in different units.

Coefficient of Variation (C.V.)

It is now clear that the standard deviation, as a measure of dispersion, gives us an idea about the extent to which observations are scattered around their mean. *Thus, two or more distributions having the same mean can be compared directly for their variability with the help of corresponding standard deviations.* Now the following two situations may arise:

(a) *When two or more distributions having unequal means are to be compared in respect of their variability.*

(b) When two or more distributions having observations expressed in different units of measurements are to be compared in respect of their scatteredness or variability.

For making comparisons in the above two situations, we use a relative measure of dispersion, called **coefficient of variation (C.V.) The coefficient of variation (C.V.)** is defined as:

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100 \qquad \text{...(25)}$$

Example 32: *A sample of 5 items was taken from the output of a factory. The length and weight of 5 items are given below*:

Length (inches)	5	6	7	9	12
Weight (ounces)	13	15	18	19	20

State which of the two characteristics of the two items is more variable.

[B.C.A. (Garhwal) 2007]

Solution: The two characteristics of the items, are length and weight expressed in inches and ounces respectively. Since the units of the two characteristics are different, their variability can be compared with the help of coefficients of variation which is computed as under:

Computation of Coefficient of Variation

Length (inches)			**Weight (ounces)**		
X	$d_X = (X-7)$	d_X^2	Y	$d_Y = (Y-18)$	d_Y^2
5	– 2	4	13	– 5	25
6	– 1	1	15	– 3	9
7	0	0	18	0	0
9	2	4	19	1	1
12	5	25	20	2	4
$N = 5$	$\Sigma d_x = 4$	$\Sigma d_x^2 = 34$	$N = 5$	$\Sigma d_y = -5$	$\Sigma d_y^2 = 39$

On comparing the two C.V.'s we can say that the length characteristic is more variable than weight.

Mean $= \overline{X} = A + \frac{\Sigma d_X}{N}$ $= 7 + \frac{4}{5} =$ **7.8 inches.**	**Mean** $= \overline{Y} = A + \frac{\Sigma d_Y}{N}$ $= 18 + \left(\frac{-5}{5}\right) =$ **17.0 ounces**
S.D. $= \sigma_X = \sqrt{\left[\frac{\Sigma d_X^2}{N} - \left(\frac{\Sigma d_X}{N}\right)^2\right]}$	**S.D.** $= \sigma_Y = \sqrt{\left[\frac{\Sigma d_Y^2}{N} - \left(\frac{\Sigma d_y}{N}\right)^2\right]}$
$= \sqrt{\left[\frac{34}{5} - \left(\frac{4}{5}\right)^2\right]}$	$= \sqrt{\left[\frac{39}{5} - \left(\frac{-5}{5}\right)^2\right]}$
$= \sqrt{[6.8 - 0.64]} = \sqrt{[6.14]}$	$= \sqrt{[7.8 - 1]} = \sqrt{(6.8)}$
= **2.48 inches.**	= **2.61 ounces.**
Co-efficient of variation C. V. $= \frac{\sigma_X}{\overline{X}} \times 100 = \frac{2.48}{7.80} \times 100$ $=$ **31.79%**	Coefficient of variation C.V. $= \frac{\sigma_Y}{\overline{Y}} \times 100 = \frac{2.61}{17.00} \times 100$ $=$ **15.35%**

On comparing the two C.V. and we can say that the length characteristic is more variable than weight.

Example 33: *Below are the number of runs scored by two batsmen in eight matches.*

Batsman A	27	16	39	45	101	80	40	52
Batsman B	0	100	80	5	60	40	10	121

Who is better run scorer ? Also find which of the two batsman is more consistent in scoring ?

[B.C.A. (Agra) 2004]

Solution: Since the average number of runs scored by the two batsmen are different, as such the coefficient of variations will be compared for observing consistency.

Computation of Coefficient of Variation

Batsman A			Batsman B		
X	$d_X = (X - 48)$	d_X^2	Y	$d_Y = (Y - 50)$	d_Y^2
27	– 21	441	0	– 50	2500
16	– 32	1024	100	+ 50	2500
39	– 9	81	80	+ 30	900
45	– 3	9	5	– 45	2025
101	+ 53	2809	60	+ 10	100
80	+ 32	1024	40	– 10	100
40	– 8	64	10	– 40	1600
52	+ 4	16	121	+ 71	5041
$N = 8$	$\Sigma d_x = 16$	$\Sigma d_x^2 = 5468$	$N = 8$	$\Sigma d_y = 16$	$\Sigma d_y^2 = 14766$

For *X*-Series	For *Y*-Series
$\textbf{Mean} = \overline{X} = A + \frac{\Sigma d_X}{N}$ $= 48 + \frac{16}{8} = \mathbf{50.0}$	$\textbf{Mean} = \overline{Y} = A + \frac{\Sigma d_Y}{N}$ $= 50 + \frac{16}{8} = \mathbf{52.0}$
$\text{S.D.} = \sigma_X = \sqrt{\left[\frac{\Sigma d_X^2}{N} - \left(\frac{\Sigma d_X}{N}\right)^2\right]}$ $= \sqrt{\left[\frac{5468}{8} - \left(\frac{16}{8}\right)^2\right]}$ $= \sqrt{(683.5 - 4)} = \sqrt{(679.5)}$ $= \mathbf{26.07}$	$\text{S.D.} = \sigma_Y = \sqrt{\left[\frac{\Sigma d_Y^2}{N} - \left(\frac{\Sigma d_Y}{N}\right)^2\right]}$ $= \sqrt{\left[\frac{14766}{8} - \left(\frac{16}{8}\right)^2\right]}$ $= \sqrt{[1845.75 - 4]}$ $= \sqrt{(1841.75)} = \mathbf{42.92}$
$\text{C.V.} = \frac{\sigma_X}{\overline{X}} \times 100 = \frac{26.07}{50.00} \times 100$ $= \mathbf{52.14\%}$	$\text{C. V.} = \frac{\sigma_Y}{\overline{Y}} \times 100 = \frac{42.92}{52.92} \times 100$ $= \mathbf{82.54\%}$

Comparing the means of runs, we find that the batsman B is better in scoring. But since, the C.V. for *X*-series is less than that of *Y*-series, thus batsman *A* is more consistent.

Example 34: *Compute coefficient of variation for factory A and B from the following:*

Daily Wages (₹)	12	15	17	22	25	30
Workers in Factory A	5	10	20	35	12	8
Workers in Factory B	15	25	30	35	21	19

Find: (a) Which factory pays higher average daily wages?

(b) In which factory, wages are more variable.

Solution: **Computation of Coefficient of Variation**

		Factory A			Factory B		
Wages X	$d_X =$ $(X - 17)$	No. of Workers f_1	$f_1 d_X$	$f_1 d_X^2$	No. of Workers f_2	$f_2 d_X$	$f_2 d_X^2$
12	– 5	5	– 25	125	15	– 75	375
15	– 2	10	– 20	40	25	– 50	100
17	0	20	0	0	30	0	0
22	5	35	175	875	35	175	875
25	8	12	96	768	21	168	1344
30	13	8	104	1352	19	247	3211
		$N_1 = 90$	$\Sigma f_1 d_X$ $= 330$	$\Sigma f_1 d_X^2$ $= 3160$	$N_2 = 145$	$\Sigma f_2 d_X$ $= 465$	$\Sigma f_2 d_X^2$ $= 5905$

For Factory A	For Factory B
Mean $= \bar{X} = A + \frac{\Sigma f_1 d_X}{N_1}$ $= 17 + \frac{330}{90} = 17 + 3.67$ = ₹ 20.67	**Mean** $= \bar{Y} = A + \frac{\Sigma f_2 d_X}{N_2}$ $= 17 + \frac{465}{145} = 17 + 3.21$ = **₹ 20.21**
S.D. $= \sigma_X = \sqrt{\left[\frac{\Sigma f_1 d_X^2}{N_1} - \left(\frac{\Sigma f_1 d_X}{N_1}\right)^2\right]}$ $= \sqrt{\left[\frac{3160}{90} - \left(\frac{330}{90}\right)^2\right]}$	**S.D.** $= \sigma_Y = \sqrt{\left[\frac{\Sigma f_2 d_X^2}{N_2} - \left(\frac{\Sigma f_2 d_X}{N_2}\right)^2\right]}$ $= \sqrt{\left[\frac{5905}{145} - \left(\frac{465}{145}\right)^2\right]}$
$= \sqrt{[35.11 - 13.44]}$ $= \sqrt{[20.67]} =$ **₹ 4.66**	$= \sqrt{[40.72 - 10.30]}$ $= \sqrt{[30.42]} =$ **₹ 5.52**
C.V. $= \frac{\sigma_X}{\bar{X}} \times 100 = \frac{4.66}{20.67} \times 100$ $=$ **22.54%**	**C.V.** $= \frac{\sigma_Y}{\bar{Y}} \times 100 = \frac{5.52}{20.21} \times 100$ $=$ **27.31%**

(a) Comparing mean wages, we can say that factory *A* pays higher average daily wages.

(b) Comparing the C.V.'s, we find that the wages are more variable in factory *B*.

Example 35: *In the following table, distribution of students is shown according to their weights in kg. Find the coefficient for each series. Which series show greater variation in weights ?*

Weight (Kg.)	20–30	30–40	40–50	50–60	60–70
Class A	7	10	20	18	7
Class B	5	9	21	15	6

[B.C.A. (Delhi) 2004]

Solution: **Computation of Coefficient of Variation**

Weight Classes	**Mid-Value** X	$d' = \frac{X-45}{10}$	**Class A**			**Class B**		
			f_1	$f_1 d'$	$f_1 d'^2$	f_2	$f_2 d'$	$f_2 d'^2$
20–30	25	– 2	7	– 14	28	5	– 10	20
30–40	35	– 1	10	– 10	10	9	– 9	9
40–50	45	0	20	0	0	21	0	0
50–60	55	1	18	18	18	15	15	15
60–70	65	2	7	14	28	6	12	24
			$N = 62$	$\Sigma f_1 d' = 8$	$\Sigma f_2 d'^2 = 84$	$N_2 = 56$	$\Sigma f_2 d' = 8$	$\Sigma f_2 d'^2 = 68$

For Class A	**For Class B**
Here $A = 45$ and $h = 10$ **Mean** $= \overline{X} = A + h\,\frac{\Sigma f_1 d'}{N_1}$ $= 45 + 10\,.\,\frac{8}{62} = 45 + 1.25$ $=$ **46.25 kg.**	Here also $A = 45$ and $h = 10$ **Mean** $= \overline{Y} = A + h\,\frac{\Sigma f_2 d'}{N_2}$ $= 45 + 10\,.\,\frac{8}{56} = 45 + 1.43$ $=$ **46.43 kg.**
S.D. $= \sigma_X = h\sqrt{\left[\frac{\Sigma f_1 d'^2}{N_1} - \left(\frac{\Sigma f_1 d'}{N_1}\right)^2\right]}$	**S.D.** $= \sigma_Y = h\sqrt{\left[\frac{\Sigma f_2 d'^2}{N_2} - \left(\frac{\Sigma f_2 d'}{N_2}\right)^2\right]}$
$= 10\sqrt{\left[\frac{84}{62} - \left(\frac{8}{62}\right)^2\right]}$ $= 10\sqrt{[1.35 - 0.0166]}$ $= 10\sqrt{[1.3334]} = 10 \times 1.1547$ $=$ **11.55 kg.**	$= 10\sqrt{\left[\frac{68}{56} - \left(\frac{8}{56}\right)^2\right]}$ $= 10 \times \sqrt{[1.2143 - 0.0204]}$ $= 10 \times \sqrt{1.1939} = 10 \times 1.093$ $=$ **10.93 kg.**

$\text{C.V.} = \frac{\sigma_X}{\overline{X}} \times 100 = \frac{11.55}{46.25} \times 100$ $= 24.57\%$	$\text{C.V.} = \frac{\sigma_Y}{\overline{Y}} \times 100 = \frac{10.93}{46.43} \times 100$ $= 23.54\%$

On comparing the coefficients of variation for the two series, we find that class A shows greater variation in weights.

Example 36: *The means and standard deviations of two brands of light bulbs are given below. Calculate a measure of relative dispersion for the two brands and interpret the result.*

	Brand-1	*Brand-2*
Mean	850 *hours*	750 *hours*
S.D.	102 *hours*	60 *hours*

[B.C.A. (Rohilkhand) 2006, 2003]

Solution:

Brand-1	Brand-2
Given that $\overline{X}_1 = 850$ hours	Given that $\overline{X}_2 = 750$ hours
$\sigma_1 = 102$ hours	$\sigma_2 = 60$ hours
$\text{C.V.} = \frac{\sigma_1}{\overline{X}_1} \times 100 = \frac{102}{850} \times 100$	$\text{C.V.} = \frac{\sigma_2}{\overline{X}_2} \times 100 = \frac{60}{750} \times 100$
$= 12\%$	$= 8\%$

Although a comparison of average life leads to the conclusion that Brand –1 is better, yet the comparison of C.V.'s shows that Brand –2 has **less variability** *or* greater uniformity. Thus, brand–2 is supposed to be better as it is more reliable.

Example 37: *In the factories A and B the average weekly wages and standard deviation are as follows*:

Factory	*Average Weekly Wages (₹)*	*S.D. of Wages*	*No. of Workers*
A	460	50	100
B	490	40	80

Answer the following:

(i) which factory pays larger amount as weekly wage ?

(ii) which factory shows greater variability in the distribution of wages ?

(iii) what is the mean and S.D. of wages of all the workers in two factories taken together ?

[B.C.A. (Meerut) 2003, 2002]

Solution: The given data can be put as:

$$\overline{X}_1 = 460, \quad \sigma_1 = 50, \quad N_1 = 100$$

$$\overline{X}_2 = 490, \quad \sigma_2 = 40, \quad N_2 = 80$$

(i) Total weekly wages paid by factory

$$A = \Sigma X = N_1\ \overline{X}_1 = 100 \times 460 = ₹\,\mathbf{46{,}000}$$

Total weekly wages paid by factory

$$B = \Sigma X = N_2\ \overline{X}_2 = 80 \times 490 = ₹\,\mathbf{39{,}200}$$

∴ Factory *A* pays larger amount as weekly wages.

(ii) For factory *A*, the C.V. $= \dfrac{\sigma_1}{\overline{X}_1} \times 100 = \dfrac{50}{460} \times 100 = \mathbf{10.87\%}$

For factory *B*, the C.V. $= \dfrac{\sigma_2}{\overline{X}_2} \times 100 = \dfrac{40}{490} \times 100 = \mathbf{8.16\%}$

The comparison of C.V.'s reveals that factory A is more variable in respect of wage distribution.

(iii) Combined Mean

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{460 \times 100 + 490 \times 80}{100 + 80} = ₹\,\mathbf{473.33}$$

For combined S.D., we first calculate—

$$D_1 = (\overline{X}_1 - \overline{X}_{12}) = 460 - 473.33 = -13.33 \quad \therefore \quad D_1^2 = 177.69$$

$$D_2 = (\overline{X}_2 - \overline{X}_{12}) = 490 - 473.33 = +16.67 \quad \therefore \quad D_2^2 = 277.89$$

$$\sigma_1^2 = (50)^2 = 2500, \ \sigma_2^2 = (40)^2 = 1600$$

Putting these values in the formula for combined S.D., *i.e.*

$$\sigma_{12} = \sqrt{\left[\frac{N_1\,(\sigma_1^2 + D_1^2) + N_2\,(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]}$$

$$= \sqrt{\left[\frac{100\,(2500 + 177.69) + 80\,(1600 + 277.89)}{100 + 80}\right]}$$

$$= \sqrt{\left[\frac{267769 + 150231.20}{180}\right]} = \sqrt{[2322.22]} = \mathbf{48.19}$$

Example 38: *Two workers on the same job shows the following results over a long period of time :*

	Worker A	*Worker B*
Mean time of completing the job (minutes)	25	20
Variance	36	16

(i) *which worker appears to be more consistent in the time he requires to complete the job ?*

(ii) *which worker appears to be faster in completing the job ?*

[B.C.A. (Delhi) 2007, 2002]

Solution:

(i) Given that $\overline{X}_1 = 25, \overline{X}_2 = 20, \sigma_1^2 = 36, \sigma_2^2 = 16.$

$$\text{For worker } A, \text{ C.V.} = \frac{\sigma_1}{\overline{X}_1} \times 100 = \frac{6}{25} \times 100 = 24\%$$

$$\text{For worker } B, \text{ C.V.} = \frac{\sigma_2}{\overline{X}_2} \times 100 = \frac{4}{20} \times 100 = 20\%$$

The comparison of C.V.'s lead to the conclusion that worker *B* appears more consistent in the time needed for completing the job.

(ii) Since the mean time taken to complete the job is less for worker *B,* thus worker *B* appears to be faster in completing the job.

Example 39: *The following distribution shows the life of bulbs manufactured by two companies X and Y :*

Use of Bulbs (In hours)	*No. of Bulbs*	
	Company (X)	*(Company Y)*
100–200	10	10
200–300	15	40
300–400	25	25
400–500	40	15
500–600	10	10

Which company is bulb have

(i) *Higher average life*

(ii) *Greater consistency ?*

[B.C.A. (AMU) 2008, 2004; B.C.A. (MDU) Rohtak 2008; B.C.A. (Avadh) 2007, 2004]

Solution:

Calculation of $\overline{X}$ and C.V (the coefficient of variance)

Use of Bulbs (in hrs)	M.V. (m)	$A=350$ $d=X-A$	Company X			Company Y		
			f	fd	fd^2	f	fd	fd^2
100–200	150	–2	10	–20	40	10	–20	40
200–300	250	–1	15	–15	15	40	–40	40
300–400	350	0	25	0	0	25	0	0
400–500	450	1	40	40	40	15	15	15
500–600	550	2	10	20	40	10	20	40
Total			N=100	$\Sigma fd=25$	$\Sigma fd^2 = 135$	$N=100$	$\Sigma fd = -25$	$\Sigma fd^2 = 135$

(i)

Company X	Company Y
$\overline{X} = A + \frac{\Sigma fd}{N}$	$\overline{Y} = A + \frac{\Sigma fd}{N}$
$= 350 + \frac{25}{100}$	$= 350 + \frac{(-25)}{100}$
$= 350 + 0.25$	$= 350 - 0.25$
$=350.25$	$= 349.75$

Since the arithmetic mean is higher for company X. Hence bulbs of company X hence higher average life.

(ii)

Company X	Company Y
$\sigma_X = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$	$\sigma_Y = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$
$\sigma_X = \sqrt{\frac{135}{100} - \left(\frac{25}{100}\right)^2}$	$\sigma_Y = \sqrt{\frac{135}{100} - \left(\frac{-25}{100}\right)^2}$
$\sigma_X = \sqrt{1.25 - 0.0625}$	$\sigma_Y = \sqrt{1.35 - 0.0625}$
$\sigma_X = 1.135$	$\sigma_Y = 1.135$

$$C.V_X = \frac{\sigma_X}{X} \times 100 = \frac{1.135}{350.25} \times 100 = 0.324\%$$

$$C.V_Y = \frac{\sigma_Y}{Y} \times 100 = \frac{1.135}{349.75} = 0.325\%$$

Since the coefficient of variation is less for company *X*, hence company X's bulbs have greater consistency.

Example 40: *From the prices of shares of X and Y given below, state which share is more stable in value.*

X:	41	44	43	48	45	46	49	50	42	40
Y:	91	93	96	92	90	97	99	94	98	95

[B.C.A. (Kurushetra) 2011, 2006]

Solution:

X	Y	$dx = X - A$ $A = 45$	dx^2	$dy = Y - A$ $A = 95$	dy^2
41	91	–4	16	–4	16
44	93	–1	1	–2	4
43	96	–2	4	1	1
48	92	3	9	–3	9
45	90	0	0	–5	25
46	97	1	1	2	4
49	99	4	16	4	16
50	94	5	25	–1	1
42	98	–3	9	3	9
40	95	–5	25	0	0
$\Sigma X = 488$ $N = 10$	$\Sigma Y = 945$ $N = 10$	$\Sigma dx = -2$	$\Sigma dx^2 = 106$	$\Sigma dy = -5$	$\Sigma dy^2 = 85$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{448}{10} = 44.8$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{945}{10} = 94.5$$

Standard deviation of *X*-series $\sigma_X = \sqrt{\frac{\Sigma dx^2}{N} - \left(\frac{\Sigma dx}{N}\right)^2}$

Here, $N = 10, \Sigma dx = 2, \Sigma dx^2 = 106$

$$\therefore \quad \sigma_X = \sqrt{\frac{106}{10} - \left(\frac{-2}{10}\right)^2} = \sqrt{10.6 - 0.04} = 3.25$$

$$\text{Coefficient of variation of } X = \frac{\sigma_X}{\overline{X}} \times 100 = \frac{3.25}{44.8} \times 100 = 7.25$$

$$\text{Standard deviation of } Y\text{-series, } \sigma_Y = \sqrt{\frac{\Sigma dy^2}{N} - \left(\frac{\Sigma dy}{N}\right)^2}$$

$$= \sqrt{\frac{85}{10} - \left(\frac{-5}{10}\right)^2} = \sqrt{8.5 - .25} = \sqrt{8.25} = 2.87$$

$$\text{Coefficient of Variation of } Y = \frac{\sigma_Y}{\overline{Y}} \times 100$$

$$= \frac{2.87}{945} \times 100 = 304$$

Since the $C.V$ is Y data is less than $C.V$ of X data, so shares of Y-series are more stable in value.

Example 41:

(i) *Coefficient of variation of two series are* 50% *and* 69%. *Their standard deviations are* 21.2 *and* 15.6 *respectively. What are their arithmetic means ?*

(ii) *Mean of* 100 *items is* 58 *and their standard deviation is* 4 *fined the sum of squares of all items.*

[B.C.A. (Delhi) 2011; B.C.A (AMU) 2011, 2007; B.C.A. (Kurukshetra) 2010; B.C.A. (Lucknow) 2008, 2004]

Solution:

(i) $C.V = \frac{\sigma}{\overline{X}} \times 100$

for series A: we have $58 = \frac{21.2}{\overline{X}_A} \times 100 \Rightarrow \overline{X}_A = 365.6$

for series B: we have $69 = \frac{15.9}{\overline{X}_B} \times 100 \Rightarrow \overline{X}_B = 22.6$

(ii) $\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\overline{X})^2}$ or $\sigma^2 = \frac{\Sigma X^2}{N} - (\overline{X})^2$

$\Rightarrow \quad (4)^2 = \frac{\Sigma X^2}{100} - (50)^2$

$\Rightarrow \quad 1600 + 250000 = \Sigma X^2$

or $\quad \Sigma X^2 = 251600$

Example 42: *Mean and standard deviation of the following continuous series are* 31 *and* 15.94 *respectively.*

The distribution after step deviation is as follows:

d	–3	–2	–1	0	1	2	3
f	10	15	25	25	10	11	5

Determine the actual class intervals. **[B.C.A. (Delhi) 2011; B.C.A. (Bundelkhand) 2008; B.C.A. (Kanpur) 2007; B.C.A. (Agra) 2004]**

Solution:

d	***f***	***fd***	***fd*²**
–3	10	–30	90
–2	15	–30	60
–1	25	–25	25
0	25	0	0
1	10	10	10
2	11	22	44
3	5	15	45
Total	$N = 101$	$\Sigma fd^2 = -38$	$\Sigma fd^2 = 274$

Where

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

$$\Rightarrow \qquad 15.94 = \sqrt{\frac{274}{101} - \left(\frac{-38}{101}\right)^2} \times i$$

$$\Rightarrow \qquad 15.94 = \sqrt{2.713 - (0.376)^2} \times i$$

$$\Rightarrow \qquad 15.94 = \sqrt{2.713 - .141} \times i \quad \text{or} \quad 15.94 = 1.60 \times i$$

$$\Rightarrow \qquad i = \frac{15.94}{1.60} = 9.96 = 10 \text{ (Approx)}$$

Also

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i$$

$$31 = A + \left(\frac{-38}{101}\right) \times 10$$

$$\Rightarrow \qquad 31 = A - 3.76$$

$$\Rightarrow \qquad A = 34.76 \quad \text{or} \quad A = 35 \text{ (Approx)}$$

Hence, $A =$ Assume mean = 35 and class interval $(i) = 10$

Example 43: *A collar manufactures is considering the production of new style of callar to attract young man. The following statistics of neck circumference are available based upon measurement of a typical group of college students:*

Mid Values in Inches	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5
Number of Students	4	19	30	63	66	29	18	1	1

Compute standard deviation and use criterian $(X \pm 30)$ *to determine the largest and smallest size of collar he should make in order to meet the needs of practically all the customers bearing in mind that collar are worn, on an average* $3/4$ *inch longer that the necks size.* **[B.C.A. (Meerut) 2004]**

Solution: Let $A = 145 \quad \therefore \quad d^1 = \frac{(X - A)}{i},\ i = 0.5$

Mid Value (X)	f	d'	d'^2	fd^1	fd'^2
12.5	4	–4	16	–16	64
13.0	19	–3	9	–57	171
13.5	30	–2	4	–60	120
14.0	63	–1	1	–63	63
14.5	66	0	0	0	0
15.0	29	1	1	29	29
15.5	18	2	4	36	72
16.0	1	3	9	3	9
16.5	1	4	16	4	16
Total	$N = 231$			$\Sigma fd^1 = -124$	$\Sigma fd'^2 = 544$

Now $$\overline{X} = A + \frac{\Sigma fd'^1}{N} \times i = 14.5 + \frac{(-124)}{231} \times 0.5 = 14.5 - 0.268 = 14.2 \text{ inches}$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{544}{231} - \left(\frac{-124}{231}\right)^2} \times 0.5 = .719 \text{ inches.}$$

Using the criterion $\overline{X} \pm 3\sigma$

Smallest size of collar $= \overline{X} - 3\sigma + \frac{3}{4}$ inch

$= 14.232 - 3(0.719) + 0.75$

$= 14.232 - 2.157 + 0.75$

$= 12.825$ inches

Largest size of the collar $= \overline{X} + 3\sigma + \frac{3}{4}$ inch

$$= 14.232 + 3(0.719) + 0.75$$

$$= 14.232 + 2.157 + 0.75$$

$$= 17.139 \text{ inches.}$$

Example 44: *Following is the data related to two factories:*

	Factory A	*Factory B*
No. of workers	200	250
Average wage per hour	₹ 15.00	₹ 12.00
Variance	16	9

Find the following:

(i) *Which factory pays larger amount as total wages per hour.*

(ii) *Which factory is more variable ?*

(iii) *Find combined S.D.* **[B.C.A. (Garhwal) 2011]**

Solution:

(i) The total wages can be obtained by finding the value of ΣX of both factories

Factory A	**Factory B**
$\overline{X}_1 = \frac{\Sigma X_1}{N}$	$\overline{X}_2 = \frac{\Sigma X_2}{N}$
$\frac{150}{1} = \frac{\Sigma X_1}{200}$	$\frac{12}{1} = \frac{\Sigma X_2}{250}$
$\Sigma X_1 =$ ₹ 3,000	$\Sigma X_2 =$ ₹ 3,000

Both factories are paying equal amount as total wage per hour

(ii) Variability can be obtained by finding the value of $C.V.$

Factory A	**Factory B**
Variance $= \sigma_1^2 = 16$	Variance $= \sigma_2^2 = 9$
$\therefore \quad \sigma_1 = \sqrt{16} = 4$	$\therefore \quad \sigma_2 = \sqrt{9} = 3$
$C.V. = \frac{\sigma_1}{X} \times 100$ $= \frac{4}{15} \times 100 = 26.67\%$	$C.V. = \frac{\sigma_2}{\overline{Y}} \times 100$ $C.V. = \frac{3}{12} \times 100 = 25\%$

Factory A is more variable at its $C.V.$ than that of $C.V.$ of factory B

(iii) Combine standard deviation = σ_{12}

$$\therefore \qquad \sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1\, d_1^2 + N_2\, d_2^2}{N_1 + N_2}}$$

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{(200 \times 15) + (250 \times 12)}{200 + 250}$$

$$= \frac{3000 + 3000}{450} = \frac{6000}{450} = 13.3$$

Where $d_1 = \overline{X}_1 - \overline{X}_{12} = 15 - 13.3 = 1.7$

$$d_2 = \overline{X}_2 - \overline{X}_{12} = 12 - 13.3 = -1.3$$

$$\therefore \qquad \sigma_{12} = \sqrt{\frac{2000 \times 16 + 250 \times 9 + (200 \times (1.7)^2) + 250 \times (-1.3)^2}{200 + 250}}$$

$$\sigma_{12} = \sqrt{\frac{3200 + 2250 + 578 + 422.5}{450}}$$

$$= \sqrt{\frac{64505}{450}} = \sqrt{14.33} = 3.78$$

Example 45: *Calculate Karl Pearson's Coefficient of skewness from the following data:*

Marks Above:	0	10	20	30	40	50	60	70	80
No. of Students:	150	140	100	80	80	70	30	14	0

[B.C.A. (Rohilkhand) 2007, 2005]

Solution: The above data are in cumulative frequency form. Firstly, we convert these data into simple form:

Marks	f	$M.V.$(m)	A=45 d(m–A)	d'	fd'	fd'^2	$c.f.$
0–10	10	5	–40	–4	–40	160	10
10–20	40	15	–30	–3	–120	360	50
20–30	20	25	–20	–2	–40	80	70
30–40	0	35	–10	–1	0	0	70
40–50	10	45	0	0	0	0	80
50–60	40	55	10	1	40	40	120
60–70	16	65	20	2	32	64	136
70–80	14	75	30	3	42	126	150
	N=150				$\Sigma fd' = -86$	$\Sigma fd'^2 = 830$	

As this is a binomial series (*i.e.* there are two maximum frequencies), we will find coefficient of skewness by using the formula

$$\text{Coefficient of } S_K = \frac{3(\overline{X} - M)}{\sigma}$$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 45 + \frac{(-86)}{150} \times 10 = 45 - \frac{86}{15} = 39.27$$

$$\text{Median item} = \text{Size of } \frac{N}{2}\text{th item} = \text{Size of } \frac{150}{2}\text{th item} = 75\text{th item}$$

Median lies in the class interval $40 - 50$

$$M = L_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Here, $L_1 = 40;\ \frac{N}{2} = 75;\ c.f. = 70;\ f = 10;\ i = 10$

$$M = 40 + \frac{75 - 70}{10} \times 10 = 45$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{830}{150} - \left(\frac{-86}{150}\right)^2} \times 10$$

$$= \sqrt{5.53 - 0.33} \times 10 = \sqrt{5.20} \times 10 = 2.28 \times 10 = 22.8$$

$\therefore$ $\overline{X} = 39.27;\ M = 45;\ \sigma = 22.8$

$$\text{Coefficient of } S_K = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(39.27 - 45)}{22.8} = \frac{3(-5.73)}{22.8} = \frac{-17.19}{22.8} = -0.75$$

List of Formulae

Range

Individual Series	Discrete and Continuous Series
Range $= L - S$	Range $= L - S$
Coeff. of Range $= \dfrac{L - S}{L + S}$	Coeff. of Range $= \dfrac{L - S}{L + S}$ Here, L is the upper limit of the highest class and S the lower limit of the lowest class.

Quartile Deviation (Q.D.)

Individual Series	Discrete and Continuous Series
Q.D. $= \dfrac{Q_3 - Q_1}{2}$	Q.D. $= \dfrac{Q_3 - Q_1}{2}$
Coeff. of Q.D. $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$	Coeff. of Q.D. $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

Mean Deviation (M.D.)

Individual Series	Discrete and Continuous Series
M.D. $= \dfrac{1}{N} \Sigma \lvert X - \overline{X} \rvert$	M.D. $= \dfrac{1}{N} \Sigma f \lvert X - \overline{X} \rvert$
Coeff. of M.D. $= \dfrac{\text{M.D.}}{\text{Mean}} = \dfrac{\text{M.D.}}{\overline{X}}$	Coeff. of M.D. $= \dfrac{\text{M.D.}}{\text{Mean}} = \dfrac{\text{M.D.}}{\overline{X}}$

Standard Deviation (S.D.)

Individual Series	Discrete and Continuous Series
(i) Actual Mean Method:	**(i) Actual Mean Method:**
$\sigma = \sqrt{\left[\dfrac{1}{N} \Sigma (X - \overline{X})^2\right]}$	$\sigma = \sqrt{\left[\dfrac{1}{N} \Sigma\, f\, (X - \overline{X})^2\right]}$
$= \sqrt{\left[\dfrac{\Sigma X^2}{N} - \left(\dfrac{\Sigma X}{N}\right)^2\right]}$	$= \sqrt{\left[\dfrac{\Sigma f X^2}{N} - \left(\dfrac{\Sigma f X}{N}\right)^2\right]}$
$= \sqrt{\left[\dfrac{1}{N} (\Sigma X^2 - N\overline{X}^2)\right]}$	$= \sqrt{\left[\dfrac{1}{N} (\Sigma f X^2 - N\overline{X}^2)\right]}$

(ii) Short-Cut Method:	**(ii) Short-Cut Method:**
$\sigma = \sqrt{\left[\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2\right]}$ where $d = (X - A)$	$\sigma = \sqrt{\left[\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2\right]}$ where, $d = (X - A)$
(iii) Step-Deviation Method:	**(iii) Step-Deviation Method:**
$\sigma = h \times \sqrt{\left[\frac{\Sigma d'^2}{N} - \left(\frac{\Sigma d'}{N}\right)^2\right]}$ where, $d' = \left(\frac{X - A}{h}\right)$	$\sigma = h \times \sqrt{\left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right]}$ where $d' = \left(\frac{X - A}{h}\right)$

Coefficient of Variation or C.V. $= \left(\frac{\sigma}{\overline{X}}\right) \times 100$

Combined S.D. $= \sigma_{12} = \sqrt{\left[\frac{N_1\,(\sigma_1^2 + D_1^2) + N_2\,(\sigma_2^2 + D_2^2)}{N_1 + N_2}\right]}$

where, $D_1 = (\overline{X}_1 - \overline{X}_{12})$ and $D_2 = (\overline{X}_2 - \overline{X}_{12})$

Variance $= \sigma^2 = (\text{S. D.})^2.$

Percentage of Items Lying between Specific σ Times

Limits	Percentage of Items
(i) within $\overline{X} \pm 1\sigma$ or one σ limits	68.27%
(ii) within $\overline{X} \pm 2\sigma$ or two σ limits	95.45%
(iii) within $\overline{X} \pm 3\sigma$ or three σ limits	99.73%

Problem Set

1. From the following data, find range and quartile deviation. Also determine their coefficients:

Months	1	2	3	4	5	6	7	8	9	10	11	12
Sales (₹)	78	80	80	82	82	84	84	86	86	88	88	90

2. Calculate Q.D. and its coefficient from the following data:

Weight (kg.)	60	61	62	63	65	70	75	80
No. of Workers	1	3	5	7	10	3	1	1

Also find range and its coefficient.

3. Calculate range and Q.D. for the following data. Also find their coefficients.

Classes	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40
Frequency	4	5	6	10	11	9	4	1

[B.C.A. (Agra) 2004]

4. Calculate mean deviation from median of the prices given below :

Prices (₹)	210	220	225	225	225	235	240	250	270	280

Also find coefficient of mean deviation.

5. From the table given below, find mean deviation from the median and also its coefficient.

Size	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	4	8	11	15	11	7	4

6. Compare the dispersion in ages of boys and girls by using coefficient of mean deviation.

Age (in years)	16	17	18	19	20	21	22	23	24	Total
Boys	4	5	7	12	20	13	5	0	4	70
Girls	2	0	4	8	15	10	6	3	2	50

[B.C.A. (Delhi) 2005]

7. Using short-cut method and step-deviation method, obtain standard deviation of the following data:

X : 420, 440, 400, 420, 470, 480, 440, 480, 450, 500

8. Ten students in a class have obtained the following marks in statistics. Find the S.D. of marks.

Student No.	1	2	3	4	5	6	7	8	9	10
Marks	5	10	20	25	40	42	45	48	70	80

[Hint. Use short-cut method]

9. Compute arithmetic mean and standard deviation from the following data :

X	1	2	3	4	5	6	7
f	6	12	18	26	16	10	8

Also find the coefficient of variation.

10. Find S.D. and C.V. from the following data:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of Students	10	15	25	25	10	10	5

11. Obtain mean and standard deviation for the following data set:

X	10	20	30	40	50	60	70
Y	05	15	30	35	25	17	08

[B.C.A. (Meerut) 2007]

12. Define S.D. Explain its features.

[B.C.A. (Meerut) 2004]

13. Define the term dispersion. Also differentiate between absolute and relative measures of dispersion.

[B.C.A. (Meerut) 2007]

Answers

1.	Range=12 ; Q.D.=3.5 ; C.R.=0.07 ; C.Q.D.=0.04
2.	Range=20 ; Q.D.=1.5 kg ; C.R.=0.43 ; C.Q.D.=0.024
3.	Range=35 ; Q.D.=6.458 ; C.R. = 0.77, C.Q.D. = 0.33
4.	M.D.=17 ; C.M.D.=0.07
5.	M.D.=12.74 ; C.M.D.=0.367
6.	Age dispersion in the boy is greater
7.	30.33
8.	38.5
9.	$\overline{X} = 4$; $\sigma = 1.607$; C.V.= 40.17%
10.	$\sigma = 15.94$; C.V.=51.42%
11.	$\bar{x} = 40.59$; $\sigma^2 = 221.87$

❑❑❑

Chapter

6

Permutations and Combinations

6.1 Factorial Notation

Let n be a positive integer. Then, the continued product of first n natural numbers is called factorial n and is denoted by $n!$ or $\lfloor\underline{n}$.

Therefore, $\quad n! = n\,(n-1)\,(n-2)\,\ldots.\,3\,.2\,.1$

or $\quad n! = 1.2\,.3\,\ldots.\,(n-1)\,n$

Illustration: $\quad 6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

Note: When n is negative or a fraction, $n!$ is not defined.

Theorem 1:

To prove that $\quad 0! = 1$

Proof: Since $\quad \lfloor\underline{n} = n\,\lfloor\underline{n-1}$

Therefore, put $\quad n = 1$, we find

$$\lfloor\underline{1} = 1 \times \lfloor\underline{0}$$

or $\quad \lfloor\underline{0} = \dfrac{\lfloor\underline{1}}{1} = 1$

or $\quad \lfloor\underline{0} = 1$

Solved Examples

Example1: *Compute.*

(i) $\dfrac{\lfloor\underline{8}}{\lfloor\underline{4} \times \lfloor\underline{3}}$ *(ii)* $\dfrac{\lfloor\underline{30}}{\lfloor\underline{28}}$ *(iii)* L.C.M. $[\lfloor\underline{4}, \lfloor\underline{5}, \lfloor\underline{6}]$

(iv) $\dfrac{\lfloor\underline{52}}{\lfloor\underline{47}\,\lfloor\underline{5}}$ *(v)* $\dfrac{1}{\lfloor\underline{5}} + \dfrac{1}{\lfloor\underline{6}} + \dfrac{1}{\lfloor\underline{7}}$

Solution:

(i) $$\frac{\lfloor\underline{8}}{\lfloor\underline{4} \times \lfloor\underline{3}} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(4 \times 3 \times 2 \times 1) \times (3 \times 2 \times 1)} = 280$$

(ii) $$\frac{\lfloor\underline{30}}{\lfloor\underline{28}} = \frac{30 \times 29 \times \lfloor\underline{28}}{\lfloor\underline{28}} = 30 \times 29 = 870$$

(iii) We have

$$\lfloor\underline{4} = 4 \times 3 \times 2 \times 1 = 2^3 \times 3$$

$$\lfloor\underline{5} = 5 \times 4 \times 3 \times 2 \times 1 = 2^3 \times 3 \times 5$$

$$\lfloor\underline{6} = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 2^4 \times 3^2 \times 5$$

L.C.M. $[\lfloor\underline{4}, \lfloor\underline{5}, \lfloor\underline{6}] = 2^4 \times 3^2 \times 5 = 720$

(iv) $$\frac{\lfloor\underline{52}}{\lfloor\underline{47}\,\lfloor\underline{5}} = \frac{52 \times 51 \times 50 \times 49 \times 48 \times \lfloor\underline{47}}{\lfloor\underline{47} \times 5 \times 4 \times 3 \times 2 \times 1} = 52 \times 51 \times 10 \times 49 \times 2 = 25{,}98{,}960$$

(v) $$\frac{1}{\lfloor\underline{5}} + \frac{1}{\lfloor\underline{6}} + \frac{1}{\lfloor\underline{7}} = \frac{1}{\lfloor\underline{5}} + \frac{1}{6\,\lfloor\underline{5}} + \frac{1}{7 \times 6 \times \lfloor\underline{5}} = \frac{42 + 7 + 1}{7 \times 6 \times \lfloor\underline{5}} = \frac{50}{7 \times 6 \times \lfloor\underline{5}} = \frac{50}{\lfloor\underline{7}}$$

Example 2: *Convert into factorials.*

(i) 6.7.8.9 *(ii)* 2.4.6.8.12

(iii) 1.3.5.7.9.11 *(iv)* 18.20.22.26.28

Solution:

(i) $$6 \cdot 7 \cdot 8 \cdot 9 = \frac{\lfloor\underline{5} \cdot 6 \cdot 7 \cdot 8 \cdot 9}{\lfloor\underline{5}} = \frac{\lfloor\underline{9}}{\lfloor\underline{5}}$$

(ii) $$2 \cdot 4 \cdot 6 \cdot 8 \cdot 10 = 2^5 \times (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5) = 2^5 \times \lfloor\underline{5}$$

(iii) $$1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 \cdot 11 = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11}{2 \cdot 4 \cdot 6 \cdot 8 \cdot 10} = \frac{\lfloor\underline{11}}{2^5\,(1 \cdot 2 \cdot 3 \cdot 4 \cdot 5)} = \frac{\lfloor\underline{11}}{2^5\,\lfloor\underline{5}}$$

(iv) $$18 \cdot 20 \cdot 22 \cdot 24 \cdot 26 \cdot 28 \cdot = 2^6\,(9 \cdot 10 \cdot 11 \cdot 12 \cdot 13 \cdot 14)$$

$$= \frac{2^6\,\lfloor\underline{8} \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 13 \cdot 14}{\lfloor\underline{8}} = 2^6\,\frac{\lfloor\underline{14}}{\lfloor\underline{8}}$$

Example 3: *Find x, if* $\frac{1}{\underline{\lfloor 4}}+\frac{1}{\underline{\lfloor 5}}+\frac{1}{\underline{\lfloor 6}}=\frac{x}{\underline{\lfloor 7}}$

Solution: We have, $\frac{1}{\underline{\lfloor 4}}+\frac{1}{\underline{\lfloor 5}}+\frac{1}{\underline{\lfloor 6}}=\frac{x}{\underline{\lfloor 7}}$

$$\Rightarrow \quad \frac{1}{\underline{\lfloor 4}}+\frac{1}{5\times\underline{\lfloor 4}}+\frac{1}{6\times 5\times\underline{\lfloor 4}}=\frac{x}{7\times 6\times 5\times\underline{\lfloor 4}}$$

$$\Rightarrow \quad \frac{30+6+1}{6\times 5\times\underline{\lfloor 4}}=\frac{x}{7\times 6\times 5\times\underline{\lfloor 4}}$$

$$\Rightarrow \quad 7\times 37=x \quad \Rightarrow x=259.$$

Example 4: *Find n, if*

(i) $\underline{\lfloor n+1}=12\times\underline{\lfloor n-1}$ *(ii)* $\frac{1}{\underline{\lfloor 9}}+\frac{1}{\underline{\lfloor 10}}=\frac{n}{\underline{\lfloor 11}}$

Solution:

(i) $\underline{\lfloor n+1}=12\times\underline{\lfloor n-1}$

$$\Rightarrow \quad (n+1)\,n\times\underline{\lfloor n-1}=12\times\underline{\lfloor n-1}$$

$$\Rightarrow \quad (n+1)\,n=12$$

$$\Rightarrow \quad n^2+n-12=0$$

$$\Rightarrow \quad (n+4)\,(n-3)=0$$

$$\Rightarrow \quad n=3 \qquad (\because n \text{ cannot be negative})$$

(ii) $\frac{1}{\underline{\lfloor 9}}+\frac{1}{\underline{\lfloor 10}}=\frac{n}{\underline{\lfloor 11}}$

$$\Rightarrow \quad \frac{1}{\underline{\lfloor 9}}+\frac{1}{10\times\underline{\lfloor 9}}=\frac{n}{11\times 10\times\underline{\lfloor 9}}$$

$$\Rightarrow \quad 1+\frac{1}{10}=\frac{n}{11\times 10}$$

$$\Rightarrow \quad n=121$$

Example 5: *If* $\underline{\lfloor n+2}=1{,}560\,\underline{\lfloor n}$*, find n.*

Solution: We have, $\underline{\lfloor n+2}=1{,}560\,\underline{\lfloor n}$

$$\Rightarrow \quad (n+2)\,(n+1)\,\underline{\lfloor n}=1{,}560\,\underline{\lfloor n}$$

$$\Rightarrow \quad (n+2)\,(n+1)=1{,}560$$

$$\Rightarrow \quad (n+2)\,(n+1)=1{,}560$$

$$\Rightarrow \quad n^2+3n+2-1{,}560=0$$

$$\Rightarrow \quad n^2+3n-1{,}558=0$$

$$\Rightarrow \qquad n = \frac{-3 \pm \sqrt{9 - 4\,(1)\,(-1{,}558)}}{2}$$

$$\Rightarrow \qquad n = \frac{-3 \pm \sqrt{6{,}241}}{2} = \frac{-3 \pm 79}{2} = 38, -41.$$

Since n can not be negative so, $n = 38$

Example 6: *Prove that* $\dfrac{\lfloor\underline{2n}}{\lfloor\underline{n}} = 1 \cdot 3 \cdot 5 \,\ldots.\, (2n-1)\, 2^n.$

Solution: We have,

$$\frac{\lfloor\underline{2n}}{\lfloor\underline{n}} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \,\ldots.\, (2n-1)\, 2n}{\lfloor\underline{n}}$$

$$= \frac{[1 \cdot 3 \cdot 5 \,\ldots.\, (2n-1)]\, [2 \cdot 4 \cdot 6 \,\ldots.\, 2n]}{\lfloor\underline{n}}$$

$$= \frac{(1 \cdot 3 \cdot 5 \,\ldots.\, 2n-1)\, (2 \times 1) \cdot (2 \times 2) \cdot (2 \times 3) \,\ldots.\, (2 \times n)}{\lfloor\underline{n}}$$

$$= \frac{1 \cdot 3 \cdot 5 \,\ldots.\, (2n-1)\, 2^n\, (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \,\ldots.\, n)}{\lfloor\underline{n}}$$

$$= \frac{1 \cdot 3 \cdot 5 \,\ldots.\, (2n-1)\, 2^n\, \lfloor\underline{n}}{\lfloor\underline{n}}$$

$$= 1 \cdot 3 \cdot 5 \,\ldots..\, (2n-1)\, 2^n.$$

Example 7: *If* $\dfrac{\lfloor\underline{n}}{\lfloor\underline{2}\ \lfloor\underline{n-2}}$ *and* $\dfrac{\lfloor\underline{n}}{\lfloor\underline{4}\ \lfloor\underline{n-4}}$ *are in the ratio* $2:1$, *find the value of n.*

Solution: We have,

$$\frac{\lfloor\underline{n}}{\lfloor\underline{2}\ \lfloor\underline{n-2}} : \frac{\lfloor\underline{n}}{\lfloor\underline{4}\ \lfloor\underline{n-4}} = 2:1$$

$$\Rightarrow \qquad \frac{\lfloor\underline{n}}{\lfloor\underline{2}\ \lfloor\underline{n-2}} \times \frac{\lfloor\underline{4}\ \lfloor\underline{n-4}}{\lfloor\underline{n}} = \frac{2}{1}$$

$$\Rightarrow \qquad \frac{4 \times 3\ \lfloor\underline{n-4}}{(n-2)\,(n-3)\ \lfloor\underline{n-4}} = \frac{2}{1}$$

$$\Rightarrow \qquad (n-2)\,(n-3) = 6$$

$$\Rightarrow \qquad n^2 - 5n + 6 = 6$$

$$\Rightarrow \qquad n^2 - 5n = 0$$

$$\Rightarrow \qquad n\,(n-5) = 0$$

$\Rightarrow$ either $\quad n = 0 \quad$ or $\quad n = 5$

But for $n = 0$, $\lfloor\underline{n-2}$ and $\lfloor\underline{n-4}$ are meaningless.

Hence, the required value of n is 5.

Example 8: *Prove that* $\lfloor\underline{n}\,(n+2) = \lfloor\underline{n} + \lfloor\underline{n+1}$

Solution: R.H.S. $= \lfloor\underline{n} + \lfloor\underline{n+1}$

$$= \lfloor\underline{n} + (n+1)\,\lfloor\underline{n}$$

$$= \lfloor\underline{n}\,(n+2) = \text{L.H.S.}$$

Exercise

1. Compute (i) $\dfrac{\lfloor\underline{32}}{\lfloor\underline{29}}$ (ii) $\dfrac{\lfloor\underline{12} - \lfloor\underline{10}}{\lfloor\underline{9}}$ (iii) $\dfrac{\lfloor\underline{16}}{\lfloor\underline{13}\,\lfloor\underline{3}}$.

2. Find L.C.M. $[\lfloor\underline{3}, \lfloor\underline{6}, \lfloor\underline{8}]$.

3. If $\dfrac{\lfloor\underline{2n}}{\lfloor\underline{3}\,\lfloor\underline{2n-3}} : \dfrac{\lfloor\underline{n}}{\lfloor\underline{2}\,\lfloor\underline{n-2}} = \dfrac{44}{3}$, find n.

4. Find x if $\dfrac{1}{\lfloor\underline{4}} + \dfrac{1}{\lfloor\underline{5}} = \dfrac{x}{\lfloor\underline{6}}$.

5. Find n, if

   (i) $\lfloor\underline{n+2} = 60 \times \lfloor\underline{n-1}$ (ii) $\lfloor\underline{n+2} = 2550 \times \lfloor\underline{n}$

6. Prove that,

   (i) $\dfrac{\lfloor\underline{n}}{\lfloor\underline{r}} = n\,(n-1)\,(n-2) \ldots. (r+1)$

   (ii) $(n-r+1) \cdot \dfrac{\lfloor\underline{n}}{\lfloor\underline{n-r+1}} = \dfrac{\lfloor\underline{n}}{\lfloor\underline{n-r}}$

   (iii) $\dfrac{\lfloor\underline{n}}{\lfloor\underline{r}\,\lfloor\underline{n-1}} + \dfrac{\lfloor\underline{n}}{\lfloor\underline{r-1}\,\lfloor\underline{n-r+1}} = \dfrac{\lfloor\underline{n+1}}{\lfloor\underline{r}\,\lfloor\underline{n-r+1}}$

7. If $\dfrac{\lfloor\underline{n}}{2\,\lfloor\underline{(n-2)}}$ and $\dfrac{\lfloor\underline{n}}{\lfloor\underline{4}\,\lfloor\underline{n-4}}$ are in the ratio $2:1$, find n.

Answers

1.	(i) 29,760 (ii) 1,310 (iii) 560			2.	40,320	3.	$n=6$
4.	36	5.	(i) 3 (ii) 49	7.	$n=5$		

6.2 Permutation

Permutation: The different arrangement which can be made out of a given number of things by taking some or all at a time, are called **Permutations**.

For example, consider arranging the digits 1, 2 and 3. The possible arrangements as follows:

$$123, 132, 231, 213, 312, 321.$$

These are 6 permutations. Here the same three digits 1, 2 and 3 have been used but the number changes when the order of digits is changed.

Thus, forming numbers with given digits means arranging the digits and hence it is the problem of permutation.

Notations: Let r and n be positive integers such that

$$1 \le r \le n.$$

Then, the number of all permutations of n things taking r at a time is denoted by

$$P(n, r) \quad \text{or} \quad {}^nP_r$$

Theorem 2: Let $1 \le r \le n$. Then the number of all permutations of n dissimilar things taken r at a time is given by

$${}^nP_r = n(n-1)(n-2)\ldots.(n-r+1) = \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}}.$$

Proof: We have

$$\text{L.H.S.} = {}^nP_r = n(n-1)(n-2)\ldots..(n-r+1)$$

$$= \frac{n(n-1)(n-2)\ldots(n-r+1)\lfloor\underline{n-r}}{\lfloor\underline{n-r}}$$

$$= \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}} = \text{R.H.S.}$$

Theorem 3: The number of all permutations of n different things taken all at a time is given by ${}^nP_n = \lfloor\underline{n}$.

Proof: We have,

$${}^nP_r = \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}}$$

Put $r = n$, we find

$${}^nP_n = \frac{\lfloor\underline{n}}{\lfloor\underline{n-n}} = \frac{\lfloor\underline{n}}{\lfloor\underline{0}} = \frac{\lfloor\underline{n}}{1} = \lfloor\underline{n}.$$

Theorem 4: Prove that $\lfloor\underline{0} = 1$.

Proof: We have,

$${}^nP_r = \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}} \qquad [\text{put } n = r]$$

$$\Rightarrow \qquad {}^nP_n = \frac{\lfloor\underline{n}}{\lfloor\underline{0}} \qquad ({}^nP_n = \lfloor\underline{n})$$

$$\Rightarrow \qquad \lfloor\underline{n} = \frac{\lfloor\underline{n}}{\lfloor\underline{0}}$$

$$\Rightarrow \qquad \lfloor\underline{0} = \frac{\lfloor\underline{n}}{\lfloor\underline{n}} = 1.$$

Example 9: *Evaluate the following:*

(i) ${}^{12}P_4$ *(ii)* ${}^{6}P_4$

(iii) ${}^{8}P_8$ *(iv)* ${}^{30}P_2$

Solution:

(i) $${}^{12}P_4 = \frac{\lfloor\underline{12}}{\lfloor\underline{12-4}} = \frac{\lfloor\underline{12}}{\lfloor\underline{8}} = \frac{12\times 11\times 10\times 9\times \lfloor\underline{8}}{\lfloor\underline{8}} = 11{,}880$$

(ii) $${}^{6}P_4 = \frac{\lfloor\underline{6}}{\lfloor\underline{6-4}} = \frac{\lfloor\underline{6}}{\lfloor\underline{2}} = \frac{6\times 5\times 4\times 3\times 2}{2} = 360$$

or $${}^{6}P_4 = 6\times 5\times 4\times 3 = 360$$

(iii) $${}^{8}P_8 = \lfloor\underline{8} = 8\times 7\times 6\times 5\times 4\times 3\times 2\times 1 = 40{,}320$$

(iv) $${}^{30}P_2 = \frac{\lfloor\underline{30}}{\lfloor\underline{30-2}} = \frac{\lfloor\underline{30}}{\lfloor\underline{28}} = \frac{30\times 29\times \lfloor\underline{28}}{\lfloor\underline{28}} = 870$$

Example 10: *If* ${}^{n}P_4 = 20 \times {}^{n}P_2$, *find n.*

Solution: We have, ${}^{n}P_4 = 20 \times {}^{n}P_2$

$$\Rightarrow \quad n(n-1)(n-2)(n-3) = 20\, n(n-1)$$

$$\Rightarrow \quad (n-2)(n-3) = 20$$

$$\Rightarrow \quad n^2 - 5n - 14 = 0$$

$$\Rightarrow \quad (n-7)(n+2) = 0$$

$$\Rightarrow \quad n = 7 \qquad (\because n \neq -2)$$

Example 11: *If* ${}^{20}P_r = 6{,}840$, *find r.*

Solution: We have, ${}^{20}P_r = 6{,}840$

$$\Rightarrow \quad \frac{\lfloor\underline{20}}{\lfloor\underline{20-r}} = 6{,}840$$

$$\Rightarrow \quad \frac{\lfloor\underline{20}}{\lfloor\underline{20-r}} = \frac{20 \times 19 \times 18 \times \lfloor\underline{17}}{\lfloor\underline{17}}$$

$$\Rightarrow \quad \frac{\lfloor\underline{20}}{\lfloor\underline{20-r}} = \frac{\lfloor\underline{20}}{\lfloor\underline{17}}$$

$$\Rightarrow \quad 20 - r = 17 \quad \text{or} \quad r = 3$$

Example 12: ${}^{n+5}P_{n+1} = \frac{11(n-1)}{2} \times {}^{n+3}P_n$, *find n.*

Solution: We have,

$${}^{n+5}P_{n+1} = \frac{11(n-1)}{2} \times {}^{(n+3)}P_n$$

$$\Rightarrow \quad \frac{\lfloor\underline{n+5}}{\lfloor\underline{(n+5)-(n+1)}} = \frac{11(n-1)}{2}\, \frac{\lfloor\underline{n+3}}{\lfloor\underline{n+3-n}}$$

$$\Rightarrow \quad \frac{(n+5)(n+4)\, \lfloor\underline{n+3}}{4 \times \lfloor\underline{3}} = \frac{11(n-1)}{2} \times \frac{\lfloor\underline{n+3}}{\lfloor\underline{3}}$$

$$\Rightarrow \quad (n+5)(n+4) = 22(n-1)$$

$$\Rightarrow \quad n^2 + 9n + 20 = 22n - 22$$

$$\Rightarrow \quad n^2 - 13n + 42 = 0$$

$$\Rightarrow \quad (n-6)(n-7) = 0$$

$$\Rightarrow \quad n = 6, 7.$$

Example 13: *If* ${}^{n}P_4 = 2 \times {}^{5}P_3$, *find n.*

Solution: ${}^{n}P_4 = 2 \times {}^{5}P_3$

$$\Rightarrow \quad n(n-1)(n-2)(n-3) = 2 \times 5 \times 4 \times 3$$

$$\Rightarrow \quad n(n-1)(n-2)(n-3) = 120$$

$$\Rightarrow \quad (n^2 - 3n)(n^2 - 3n + 2) = 120$$

$$\Rightarrow \quad m(m+2) = 120 \text{ where } m = n^2 - 3n$$

$$\Rightarrow \quad m^2 + 2m - 120 = 0$$

$$\Rightarrow \quad (m+12)(m-10) = 0$$

$$\Rightarrow \quad m = -12, \text{ or } m = 10$$

$$\Rightarrow \quad n^2 - 3n = -12 \text{ or } n^2 - 3n = 10$$

$$\Rightarrow \quad n^2 - 3n + 12 = 0 \text{ or } n^2 - 3n - 10 = 0$$

$$\Rightarrow \quad n = \frac{3 \pm \sqrt{9 - 48}}{2} \text{ or } (n-5)(n+2) = 0$$

$$\Rightarrow \quad n = \frac{3 \pm i\sqrt{39}}{2} \quad \text{or} \quad n = 5 \quad \text{or} \quad n = -2$$

$$\Rightarrow \quad n = 5 \qquad \text{[neglecting negative and imaginary value]}$$

Example 14: *Find n if* ${}^{9}P_5 + 5 \cdot {}^{9}P_4 = {}^{10}P_n$.

Solution: We have, ${}^{9}P_5 + 5 \cdot {}^{9}P_4 = {}^{10}P_n$

$$\Rightarrow \quad \frac{\lfloor\underline{9}}{\lfloor\underline{9-5}} + 5 \cdot \frac{\lfloor\underline{9}}{\lfloor\underline{9-4}} = \frac{\lfloor\underline{10}}{\lfloor\underline{10-n}}$$

$$\Rightarrow \quad \frac{\lfloor\underline{9}}{\lfloor\underline{4}} + 5 \cdot \frac{\lfloor\underline{9}}{\lfloor\underline{5}} = \frac{\lfloor\underline{10}}{\lfloor\underline{10-n}}$$

$$\Rightarrow \quad \frac{\lfloor\underline{9}}{\lfloor\underline{4}} + \frac{\lfloor\underline{9}}{\lfloor\underline{4}} = \frac{\lfloor\underline{10}}{\lfloor\underline{10-n}}$$

$$\Rightarrow \quad 2 \times \frac{\lfloor\underline{9}}{\lfloor\underline{4}} = \frac{10 \times \lfloor\underline{9}}{\lfloor\underline{(10-n)}}$$

$$\Rightarrow \quad 5 \times \lfloor\underline{4} = \lfloor\underline{10-n}$$

$$\Rightarrow \qquad \underline{\lfloor 10-n} = \underline{\lfloor 5}$$

$$\Rightarrow \qquad 10-n=5$$

$$\Rightarrow \qquad n=5.$$

Example 15: *Prove that*

$$^{n}P_r = {}^{n-1}P_r + r \cdot {}^{(n-1)}P_{r-1}$$

Solution: R.H.S. $= {}^{(n-1)}P_r + r \cdot {}^{(n-1)}P_{r-1}$

$$= \frac{\underline{\lfloor n-1}}{\underline{\lfloor n-1-r}} + r \cdot \frac{\underline{\lfloor n-1}}{\underline{\lfloor (n-1)-(r-1)}}$$

$$= \underline{\lfloor n-1}\left[\frac{1}{\underline{\lfloor n-r-1}} + \frac{r}{\underline{\lfloor n-r}}\right]$$

$$= \underline{\lfloor n-1}\left[\frac{1}{\underline{\lfloor n-r-1}} + \frac{r}{(n-r)\,\underline{\lfloor n-r-1}}\right]$$

$$= \underline{\lfloor n-1}\left[\frac{n-r+r}{(n-r)\,\underline{\lfloor n-r-1}}\right]$$

$$= \frac{n\,\underline{\lfloor n-1}}{(n-r)\,\underline{\lfloor n-r-1}} = \frac{\underline{\lfloor n}}{\underline{\lfloor n-r}} = {}^{n}P_r = \text{L. H. S.}$$

Example 16: *Prove that*

(i) $^{n}P_n = {}^{n}P_{(n-1)}$ *(ii)* $^{n}P_r = n \cdot {}^{n-1}P_{(r-1)}$

Solution:

(i) $$^{n}P_{n-1} = \frac{\underline{\lfloor n}}{\underline{\lfloor n-(n-1)}} = \frac{\underline{\lfloor n}}{\underline{\lfloor 1}} = \underline{\lfloor n} = {}^{n}P_n$$

(ii) $$^{n}P_r = \frac{\underline{\lfloor n}}{\underline{\lfloor n-r}} = \frac{n\,\underline{\lfloor n-1}}{\underline{\lfloor (n-1)-(r-1)}} = n \cdot {}^{(n-1)}P_{r-1}$$

Example 17: *If* $^{56}P_{r+6} : {}^{54}P_{r+3} = 30{,}800 : 1$, *find r.*

Solution: We have

$$^{56}P_{r+6} = \frac{\underline{\lfloor 56}}{\underline{\lfloor 56-(r+6)}} = \frac{\underline{\lfloor 56}}{\underline{\lfloor 50-r}}$$

and $$^{54}P_{r+3} = \frac{\underline{\lfloor 54}}{\underline{\lfloor 51-r}}$$

$$\therefore \qquad {}^{56}P_{r+6} : {}^{54}P_{r+3} = \frac{\underline{|56}\,/\,\underline{|50-r}}{\underline{|54}\,/\,\underline{|51-r}} = \frac{30,800}{1}$$

$$\Rightarrow \qquad 56 \times 55 \times (51-r) = 30,800$$

$$\Rightarrow \qquad 51 - r = 10 \Rightarrow r = 41$$

Example 18: Prove that

$${}^{8}P_{5} = 8.\,{}^{7}P_{4}$$

Solution: We have

$$\text{L.H.S.} = {}^{8}P_{5} = \frac{\underline{|8}}{\underline{|8-5}} = \frac{\underline{|8}}{\underline{|3}}$$

$$= \frac{8 \times 7 \times 6 \times 5 \times 4 \times \underline{|3}}{\underline{|3}} = 6,720$$

$$\text{R.H.S.} = 8\,{}^{7}P_{4} = 8.\frac{\underline{|7}}{\underline{|7-4}}$$

$$= 8 \times \frac{7 \times 6 \times 5 \times 4 \times \underline{|3}}{\underline{|3}}$$

$$= 6,720$$

$$\Rightarrow \qquad \text{L.H.S.} = \text{R.H.S.}$$

Example 19: If ${}^{111}P_{93} = {}^{110}P_{93} + x.{}^{110}P_{92}$ *then find the value of* ***x***.

[B.C.A. (Agra) 2008]

Solution: Here, we have

$${}^{111}P_{93} = {}^{110}P_{93} + x.{}^{110}P_{92} \qquad ..(1)$$

We know that

$${}^{n}P_{r} = {}^{n-1}P_{r} + r.{}^{n-1}P_{r-1} \qquad ...(2)$$

Comparing (1) and (2), we get

$$x = r = 93$$

Example 20: *If* ${}^{n}P_{100} = 103.\,{}^{102}P_{99}$ *then find the value of* n.

Solution: Here, we have

$${}^{n}P_{100} = 103.{}^{102}P_{99} \qquad ...(1)$$

We know that

$${}^{n}P_{r} = n.{}^{n-1}P_{r-1} \qquad ..(2)$$

comparing (1) and (2), we get

$$n = 103$$

Example 21: If $^{x}P_5 = 5 \times 4 \times 3 \times 2 \times 1$ *find x.*

[B.C.A. (A.M.U.) 2010]

Solution: Here, we have

$$^{x}P_5 = 5 \times 4 \times 3 \times 2 \times 1 \qquad ...(1)$$

We know that

$$^{n}P_n = \underline{|n} = n \times (n-1) \times \ldots \times 3 \times 2 \times 1 \qquad ...(2)$$

From (2), we have

$$^{5}P_5 = \underline{|5} = 5 \times 4 \times 3 \times 2 \times 1 \qquad ...(3)$$

Comparing (1) and (3), we get

$$x = 5$$

Example 22: *Each of 2 women and 3 men is to occupy one chair out of 8 chairs, each of which is numbered from 1 to 8.*

First, women are to occupy any two chairs from those numbered 1 to 4; and then the three men would occupy any three chairs out of the remaining 6 chairs. What is the maximum number of different ways in which this can be done ?

[B.C.A. (Lucknow) 2010]

Solution: First women can occupy one chair numbered from 1 to 4 in $^{4}P_1 = 4$ ways.

Second women can occupy one chair out of the remaining 3 chairs (numbered 1 to 4) in $^{3}P_1 = 3$ ways.

∴ Ways of occupying chair by women's $= 4 \times 3 = 12$

Ways of occupying the 6 chairs by 3 men's (out of the remaining 6 chairs) $= {}^{6}P_3 = 120$

∴ The maximum number of different ways of occupying chair $= 12 \times 120 = 1440$ **Ans.**

Example 23: *A question paper had 10 questions. Each question could only be answered as True (T) or False (F). Each candidate answered all the questions. Yet, no two candidates wrotes the answers in an identical sequence.*

How many different sequences of answers are possible ?

[B.C.A. (B.H.U.) 2010]

Solution: Here the questions are non-repeatable whereas the answers (T or F) are repeatable.

∴ No. of different sequences of answers $= 2^{10} = 1024$

Example 24: *A person X has four notes of rupee (₹) 1, 2, 5 and 10 denomination. The number of different sums of money she can form them is ?*

[B.C.A. (Kanpur) 2010]

Solution: The number of different sums

$$= {}^4P_2 = \frac{\lfloor\underline{4}}{\lfloor\underline{4-2}} = \frac{24}{2} = 12$$

Example 25: *In how many ways can four children be made to stand in a line such that two of them, A and B are always together ?* **[B.C.A. (Kanpur) 2007]**

Solution: Let A, B, C and D are four children out of which AB is considered as one child thus we have three child which can be arranged in a line in ${}^3P_3 = \lfloor\underline{3} = 6$ ways and the pair AB itself can be arranged in ${}^2P_2 = \lfloor\underline{2} = 2$ ways.

$\therefore$ The required no. of ways $= 6 \times 2 = 12$

Exercise

1. Prove that ${}^9P_3 + 3 \cdot {}^9P_2 = {}^{10}P_3$.
2. Prove that ${}^nP_n = 2 \cdot {}^nP_{(n-2)}$.
3. Find all permutations of 7 objects taken 3 at a time.
4. (i) If ${}^nP_4 : {}^nP_5 = 1 : 2$, find n.
   (ii) If ${}^{(n-1)}P_3 : {}^{(n+1)}P_3 = 5 : 12$, find n.
   (iii) If ${}^{2n+1}P_{n-1} : {}^{(2n-1)}P_n = 3 : 5$, find n.
5. If ${}^{11}P_r = {}^{12}P_{r-1}$, find r.
6. (i) If ${}^{2n}P_3 = 100 \times {}^nP_2$, find n.
   (ii) If $16\ {}^nP_3 = 13\ {}^{n+1}P_3$, find n.
   (iii) If ${}^nP_5 = 20 \times {}^nP_3$, find n.

Answers

3.	210	4.	(i) 6 (ii) 8 (iii) 4.
5.	9	6.	(i) 13 (ii) 15 (iii) 8

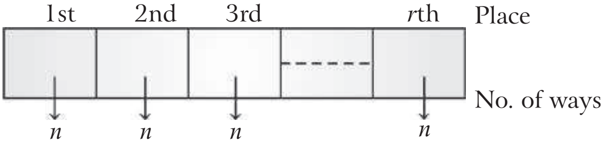
6.2.1 Permutation with Repetition

If out of n objects in a set, p objects are exactly alike of one kind, q objects exactly alike of second kind and r objects exactly alike of third kind and remaining objects are all different then the number of permutation of n objects taken all at a time is

$$= \frac{\underline{\lfloor n}}{\underline{\lfloor p}\,\underline{\lfloor q}\,\underline{\lfloor r}}$$

Theorem 5: The number of permutations of n different objects taken r at a time when each object may be replaced any number of times in each arrangement is n^r.

Proof : The number of permutations of n objects taken r at a time is the same as the number of ways of filling r places with n different objects.



Number of ways of filling first place = n,

Number of ways of filling second place = n,

(Since the object used in filling the first place can be repeated)

Number of way of filling third place = n,

....

....

....

Number of way of filling rth place = n

Therefore, by the fundamental principle of counting, the required number of permutations

$$= n \cdot n \cdot n \ldots\ldots r \text{ (times)} = n^r$$

Example 26: *How many ways are there to arrange the nine letters in the word ALLAHABAD.*

Solution: Since the word ALLAHABAD contains 4A's and 2L's, therefore, the total number of arrangement of nine letters in the word ALLAHABAD

$$= \frac{\underline{\lfloor 9}}{\underline{\lfloor 4}\,\underline{\lfloor 2}} = 7{,}560 \text{ ways.}$$

Example 27: *Find the number of different messages that can be represented by sequences of four dashes and three dots.*

Solution: The number of different message sent by 4 deshes and 3 dots

$$=\frac{\lfloor\underline{7}}{\lfloor\underline{4}\,\lfloor\underline{3}}=35$$

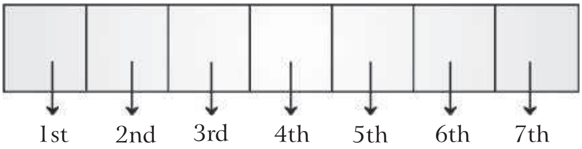
Example 28: *Find the number of ways to point 12 offices so that 3 of them are green, 2 of them pink, 2 of them yellow and the remaining one white.*

Solution: The required number of ways

$$=\frac{\lfloor\underline{12}}{\lfloor\underline{3}\,\lfloor\underline{2}\,\lfloor\underline{2}\,\lfloor\underline{5}}=1,66,320$$

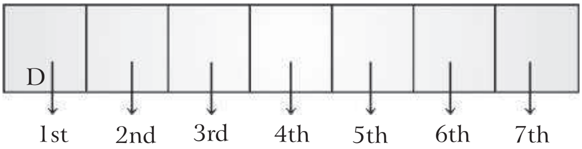
Example 29: *The letters of the word "TUESDAY" are arranged in a line each arrangement ending with letter S. How many different arrangements are possible ? How many of them start with letter D ?*

Solution: There are seven different letter in the word "TUESDAY"



when last place is filled with letter S then remaining 6 places can be filled by 6 letters in

$$^6P_6=6\,!\,\text{ways}=720\text{ ways}$$



Now, when the letter 'D' is also fixed, then remaining 5 places can be filled by 5 letters in

$$^5P_5=5\,!\ \text{ways}=120\text{ ways.}$$

Example 30: *How many signals can be made by using 6 flags of different colours when any number of them may be hoisted at a time ?*

Solution: Number of signals that can be made using 1 flag

$$={}^6P_1=6$$

Number of signals that can be made using 2 flags

$$={}^6P_2=\frac{6\,!}{(6-2)\,!}=6\times5=30$$

Number of signals that can be made using 3 flags

$$= {}^6P_3 = \frac{6!}{(6-3)!} = 6 \times 5 \times 4 = 120$$

Number of signals that can be made using 4 flags

$$= {}^6P_4 = \frac{6!}{(6-4)!} = 6 \times 5 \times 4 \times 3 = 360$$

Number of signals that can be made using 5 flags

$$= {}^6P_5 = \frac{6!}{(6-5)!} = 6 \times 5 \times 4 \times 3 \times 2 = 720$$

Number of signals that can be made using 6 flags

$$= {}^6P_6 = 6! = 720$$

Therefore the total signals that can be made

$$= 6 + 30 + 120 + 360 + 720 + 720 = 1{,}956$$

Example 31: *Find the number of 4 digit numbers that can be formed using the digits 1, 2, 3, 4 and 5 if no digit is used more than once in a number. How many of these number will be even ?*

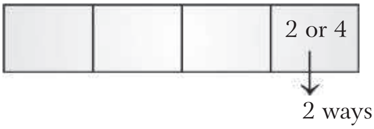
Solution:

(i)

4 places can be filled by 5-digits in ${}^5P_4$ ways

$$= \frac{5!}{(5-4)!} \text{ ways} = \frac{5!}{1!} \text{ ways} = 120 \text{ ways}$$

(ii)



For even numbers, unit's place can be filled in 2 ways

(either digit 2 or digit 4)

If units place is filled by any one digit, then remaining 3 places can be filled by remaining 4 digit in ${}^4P_3 = \frac{4!}{(4-3)!}$

$$= 4! = 24 \text{ ways}$$

Hence, total number of ways $= 2 \times 24 = 48$ ways.

Example 32: *How many 4 letter words, with or without meaning, can be formed out of the letters of the word 'LOGARITHMS' if repetition of letters is not allowed ?*

Solution: The word 'LOGARITHMS' contains different letters

∴ the number of required words

$$= \text{number of arrangements of 10 letters, taken 4 at a time}$$

$$= {}^{10}P_4$$

$$= 10 \times 9 \times 8 \times 7 = 5{,}040$$

Example 33: *It is required to seat 5 men and 4 women in a row so that the women occupy the even places. How many such arrangement are possible ?*

Solution: In a row of 9 positions, the 2nd, 4th, 6th and 8th are the even places.

These 4 places can be occupied by 4 women in

$$^4P_4 = \lfloor\underline{4} = 24 \text{ ways}$$

The remaining positions can be occupied by 5 men in

$$^5P_5 = \lfloor\underline{5} = 120 \text{ ways}$$

∴ the total number of seating arrangements $= 24 \times 120 = 2{,}880$.

Example 34: *How many words, with or without meaning, can be founded by using all letters of the word 'DELHI', using each letter exactly once ?*

Solution: The word 'DELHI' contains 5 different letters.

∴ the number of required words

$$= \text{number of arrangements of 5 letters taken all at a time}$$

$$= {}^5P_5 = \lfloor\underline{5} = 5 \times 4 \times 3 \times 2 = 120.$$

Example 35: *In an examination hall, there are four rows of chairs. Each row has 8 chairs, one behind the other. There are two classes appearing for the examination with 16 students in each class. It is desired that in each row all students belong to the same class and that no two adjacent rows are allotted to the same class. In how many way can these 32 students be seated ?*

Solution: There are 2 ways of selecting the rows for a class *i.e.*, either I and III or II and IV.

Now, 16 students of one class can be arranged in 16 chairs in

$$^{16}P_{16} = \lfloor\underline{16} \text{ ways}$$

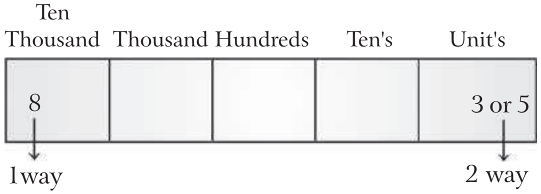
and 16 students of another class can be arranged in 16 remaining chairs in

$$^{16}P_{16} = \lfloor\underline{16} \text{ ways}$$

∴ the required number of ways $= [2 \times \lfloor\underline{16} \times \lfloor\underline{16}\ ]$

Example 36: *How many odd numbers greater than* 80,000 *can be formed using the digits* 2, 3, 4, 5 *and* 8. *If each digit is used only once in a number ?*

Solution:



Since, it is an odd number greater than 80,000, therefore unit's place can be filled by 2 ways (either by digit 3 or by digit 5) and ten thousand's can be filled by only 1 digit *i.e.*, 8.

Also other 3 places can be filled by remaining 3 digits in

$$^{3}P_{3} = 3!\,\text{ways}$$

$$\text{Total number of ways} = 2 \times 1 \times 3! = 12 \text{ (By F.P.C.).}$$

Example 37: *How many four-digit numbers (between* 1,000 *and* 10,000*) are there with distinct* 2 *of these, how many are odd numbers ?*

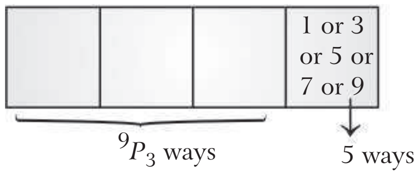
Solution:

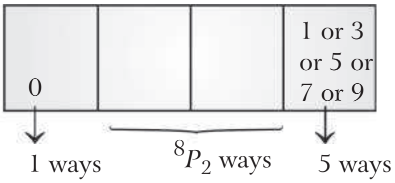
(i) Total number of arrangements of ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 taking 4 at a time is $^{10}P_4$. But these arrangements, also include those numbers which has 0 at thousand's place. Such numbers are not the four digit numbers. So, we have to exclude those three digit numbers when 0 is in thousand's place, we have to arrange remaining 9 digits by taking 3 digits at a time. The number of such arrangements is $^{9}P_3$.

Hence, the total number of four digit numbers

$$= {}^{10}P_4 - {}^{9}P_3 = 5{,}040 - 504 = 4{,}536.$$

(ii) Now, since we want the odd number, therefore unit's place can be filled in 5 (1 or 3 or 5 or 7 or 9) ways.





The required number of odd numbers = number of all 4 digited number (including 0) having an odd digit in thousand's place – number of these 4 digited odd numbers which have 0 in thousand's place

$$= 5 \times {}^{9}P_3 - 1 \times {}^{8}P_2 \times 5 = 5 \times 9 \times 8 \times 7 - 8 \times 7 \times 5 = 2{,}240.$$

Example 38: *Find the number of permutations of n different things taken r at a time each of which includes 3 particular things ?*

Solution: Given r places, 3 places by particular things can be filled in ${}^{r}P_3$ ways.

Now, we have to fill up the remaining $(r-3)$ places with the remaining $(n-3)$ things in ${}^{n-3}P_{r-3}$ ways.

Total number of permutations $= {}^{r}P_3 \times {}^{n-3}P_{r-3}$.

Example 39: *Find the number of permutations of n different things taken r at a time from which three particular things are to be excluded.*

Solution: We exclude the 3 particular things from the total number of things and form permutations of the remaining $(n-3)$ things taken r at a time.

Therefore, the required number of permutations

$$= {}^{n-3}P_r.$$

Example 40: *How many permutation can be made out of the letters of the word 'TRIANGLE' How many of these will begin with T and end with E ?*

Solution: The word TRIANGLE contains 8 different letters.

The number of all permutation made out of 8 letters

$$= {}^{8}P_8 = \lfloor\underline{8} = 8\times7\times6\times5\times4\times3\times2\times1 = 40{,}320.$$

If the word starts with the letter T and ends with letter E then, the number of ways to form the words $= 8P_6$

$$= \frac{\lfloor\underline{8}}{\lfloor\underline{8-6}} = \frac{\lfloor\underline{8}}{\lfloor\underline{2}}$$

$$= 20{,}160$$

Example 41: *In how many ways can 10 books be arranged on a shelf so that a particular pair of books shell be (i) always together (ii) never together ?*

Solution:

(i) Since a particular pair of books is always together, let us tie these two books together and consider the pair as one book. Then, we shall have to arrange 9 books on the shelf.

This can be done ${}^{9}P_9 = \lfloor\underline{9}$ ways

Now, in each of these arrangements, the two books comprising the particular pair can be arranged among themselves in $\lfloor\underline{2}$ *i.e.*, 2 ways.

$\therefore$ the required number of ways $= 2 \times (\lfloor\underline{9})$

(ii) we know that 10 books can be arranged on the shelf in

$$^{10}P_{10} = \lfloor\underline{10} \text{ ways}$$

Also, the number of ways of arranging 10 books so that a particular pair is always together $= 2 \times (\lfloor\underline{9})$

∴ the number of ways of arranging 10 books so that a particular pair is never together $= (\lfloor\underline{10} - 2\,\lfloor\underline{9})$

$$= (10 \times \lfloor\underline{9} - 2 \times \lfloor\underline{9})$$

$$= (8 \times \lfloor\underline{9})$$

Example 42: *There are 6 English, 4 Sanskrit and 5 Hindi books. In how many ways can they be arranged on a shelf so as to keep all the books of the same language together ?*

Solution: Let us make one packet for each of the books on the same language.

Now, 3 packets can be arranged in $^3P_3 = \lfloor\underline{3} = 6$ ways

6 books on English can be arranged in $\lfloor\underline{6} = 720$ ways

4 books on Sanskrit can be arranged in $\lfloor\underline{4} = 24$ ways

5 books on Hindi can be arranged in $\lfloor\underline{5} = 120$ ways

∴ the required number of ways $= 6 \times 720 \times 24 \times 120$.

Example 43: *In how many ways 9 pictures can be hung of 6 picture nails on a wall (in a row) ?*

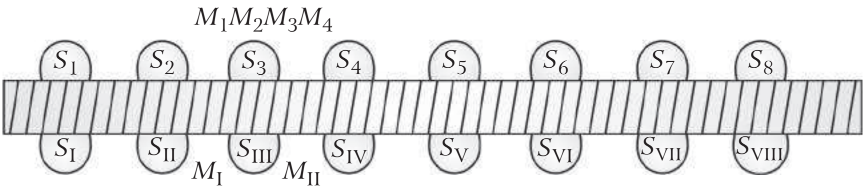
Solution: Here, it is obvious that only 6 pictures out of 9 will be hung and 3 will be left out. The number of ways in which pictures can be hung on 6 picture nails on a wall is the same as the number of arrangements of 9 things, taking 6 at a time.

Hence, the required number $= {}^9P_6$

$$= 9 \times 8 \times 7 \times 6 \times 5 \times 4 = 60{,}480.$$

Example 44: *A tea party is arranged for 16 people along the two sides of a long table with 8 chairs on each side. Four men wish to sit on one particular side and two on the other side. In how many ways can they be seated ?*

Solution: The order of seating arrangement is as below:



(i) 4 persons (M_1, M_2, M_3, M_4) are to sit one-one side in any one side in any of 8 chairs in $^8P_4$ ways.

(ii) 2 persons $(M_1$ and $M_2)$ wist to sit at other side on any of the 8 chairs. This can be in $^8P_2$ ways.

(iii) Rest 10 persons can sit on any of the remaining 10 chairs in 10 chairs in $^{10}P_{10}$ ways.

∴ Total number of ways of sitting

$$= {}^8P_4 \times {}^8P_2 \times {}^{10}P_{10} \qquad \text{(by using F.P.C.).}$$

Example 45: *In how many ways can 6 balls of different colours, namely, white, black, blue, red, green and yellow be arranged in a row in such a ways that the white and black balls are never together ?*

Solution: Let x be the number of all arrangements, each one of which contains white and black balls together.

And, let y be the number of arrangements, none of which contains white and black balls together.

Then, $x + y = {}^6P_6 = \underline{|6} = 720.$

Now, considering a white and a black ball tied together as one ball.

The number of ways of arranging the 5 balls

$$= {}^5P_5 = \underline{|5} = 120.$$

Also, these 2 balls may be arranged among themselves in $\underline{|2} = 2$ ways.

∵ $x = 2 \times 120 = 240$

So, $y = 720 - 240 = 480$

Hence, the requisite number of ways $= 480$.

Example 46: *In how many ways can the word 'PENCIL' be arranged so that N is always next to E ?*

Solution: Keeping E and N together and considering then as one letter, we have to arrange 5 letters at 5 places. This can be done in

$$^5P_5 = \underline{|5} = 120 \text{ ways.}$$

Example 47: *The principal wants to arrange 5 students on the platform such that the boy Salim occupies the second position and such that the girl Sita is always adjacent to the girl Rita. How many such arrangements are possible ?*

Solution: Let the seats be arranged as shown below:

O	S	O	O	O
I	II	III	IV	V

Keep Salim fixed at the second position.

Since, Sita and Rita are to sit together, none of the two can occupy the first seat.

This seat can be occupied by any of the remaining two students in 2 ways.

Now, two seats, namely, III , IV, V, may be occupied by Sita and Rita in 4 ways.

The remaining seat may now be occupied by the 5th student in one way only.

$\therefore$ The number of required arrangement $= 2 \times 4 \times 1 = 8.$

Example 48: *Find the number of ways in which 5 boys and 3 girls can be seated in a row so that two girls are together.*

Solution: The 5 boys may occupy 5 places in

$$^5P_5 = \lfloor\underline{5} = 120 \text{ ways.}$$

Since no. two girls are to sit together, we may arrange the 3 girls at the 6 places.

i.e., X B X B X B X B X B

Now, the 3 girls at 6 places may be seated in

$$^6P_3 = 6 \times 5 \times 4 = 120 \text{ ways.}$$

$\therefore$ The required number of ways $= 120 \times 120 = 14{,}400.$

Example 49: *How many permutations of the letters of the word 'APPLE' are there ?*

Solution: Here there are 5 objects, two of which are of same kind and the others are each of its own kind.

$\therefore$ The required number of permutations

$$= \frac{\lfloor\underline{5}}{\lfloor\underline{2} \times \lfloor\underline{1} \times \lfloor\underline{1} \times \lfloor\underline{1}} = 60.$$

Example 50: *(i) Find how many arrangements can be made with the letters of the word 'MATHEMATICS'.*

(ii) In how many of them are the vowels together ?

Solution:

(i) There are 11 letters in the word

'MATHEMATICS'

Out of these letters M occurs twice. A occurs twice, T occurs twice and the rest are all different.

$\therefore$ Required number of arrangement $= \dfrac{\lfloor\underline{11}}{\lfloor\underline{2} \times \lfloor\underline{2} \times \lfloor\underline{2}} = 49{,}89{,}600.$

(ii) The given word consists of 4 vowels namely A, E, A and I. Treating these four vowels as one letter, we have to arrange 8 letters

(MTHMTCS) + (AEAI); out of which M occurs twice, T occurs twice and the rest are all different.

$\therefore$ The number of such arrangements $= \dfrac{\lfloor\underline{8}}{(\lfloor\underline{2}) \times (\lfloor\underline{2})} = 10{,}080$

Corresponding to each such arrangement, the four vowels in which A occurs twice and the rest are all different, can be arranged amongest themselves in

$$\frac{\lfloor\underline{4}}{\lfloor\underline{2}} = 12 \text{ ways.}$$

$\therefore$ Total number of arrangements in which vowels are always together

$$= 10080 \times 12 = 1{,}20{,}960.$$

Example 51: *In the different permutations of the word 'EXAMINATION' are listed as in a dictionary, how many items are there in the list before the first word starting with E ?*

Solution: Starting with E, we have to find the number of arrangements of the remaining 10 letters, EXAMINATION, out of which there are 2I's, 2N's, 2 A's and the rest of the letters are distinct.

The number of such arrangements is

$$= \frac{\lfloor\underline{10}}{\lfloor\underline{2} \times (\lfloor\underline{2}) \times (\lfloor\underline{2})} = 4{,}53{,}600$$

Clearly, we will have to start the next word with E, so the required numbers of items $= 4{,}53{,}600$.

Example 52: *How many numbers greater than a million can be formed with digits* 2,3,0,3,4,2,3?

Solution: Any number greater than one million will contain all the seven digits.

Now, we have to arrange seven digits, out of which 2 occurs twice, 3 occurs thrice and the rest are distinct.

The number of such arrangements $= \dfrac{\lfloor\underline{7}}{(\lfloor\underline{2}) \times \lfloor\underline{3}} = 420$

These arrangements will also include those which contain 0 at the million's place. Keeping 0 fixed at the million's place, the remaining 6 digits out of which 2 occurs twice, 3 occurs thrice and the rest are distinct can be arrange in

$$= \frac{\lfloor\underline{6}}{(\lfloor\underline{2}) \times \lfloor\underline{3}} = 60 \text{ ways.}$$

$\therefore$ The number of required numbers $= 420 - 60 = 360$.

6.2.2 Circular Permutations

So far we have been considering the arrangements of objects in a line. Such permutation are known as linear permutations.

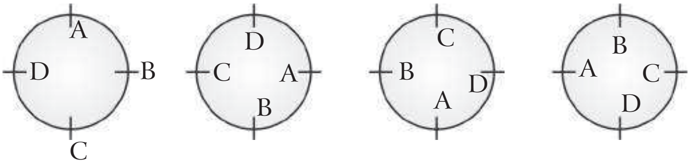
In circular permutations, what really matters is the position of an object relative to the others.

If we consider the linear permutations

ABCD, BCDA, CDAB and DABC

Then, clearly they are distinct.

Now, we arrange A, B, C, D along the circumference of a circle as shown:



If we consider the position of an object relative to others then we find that the above four arrangements are the same.

Theorem 6: The number of circular permutations of n different objects, is $\underline{\lfloor n-1}$.

Proof: Fixing the position of an object can be done in n ways, as the position of anyone of them may be fixed.

Thus, each circular permutation corresponding to n linear permutations depending upon where from we start.

Since, there are $\underline{\lfloor n}$ linear permutations, it follows that there are $\frac{\underline{\lfloor n}}{n}$

i.e., $\underline{\lfloor (n-1)}$ circular permutations.

Theorem 7: The number of ways in which n persons can be seated round a table is $\underline{\lfloor n-1}$.

Proof: Let us fix the position of one person and then arrange the remaining $(n-1)$ persons in all possible ways. Clearly, this can be done in $\underline{\lfloor (n-1)}$ ways.

Hence, the required number of ways = $\underline{\lfloor n-1}$.

Theorem 8: Show that the number of ways in which n different beads can be arranged to form a necklace is $\frac{1}{2}\underline{\lfloor (n-1)}$.

Proof: Fixed the position of one bead, the remaining $(n-1)$ beads can be arranged in $\underline{\lfloor n-1}$ ways.

In case of arranging the beads, there is no distinction between the clockwise and anticlockwise arrangements.

So, the required number of ways $=\frac{1}{2}\underline{|(n-1)}$.

Example 53: *A gentleman has 6 friends to invites. In how many ways can he send invitation cards to them, if he has 3 servants to carry the cards ?*

Solution: Two friends can be invited by the same servant. Hence, in this problem, the servant is repeatable (R) and the friend is non-repeatable (NR).

$$\text{Number of ways} = [R]^{NR} = 3^6 = 729.$$

Example 54: *A telegraph has 5 arms and each arm is capable of 4 distinct positions, including the position of rest. Find the total signals that can be made ?*

Solution: Two arms may have the same position, but same arms cannot have two positions at a time. Hence, position is repeatable (R), but arm is non-repeatable (NR)

$$\text{Number of ways} = [R]^{NR} = 4^5 = 1{,}024$$

But, in one case, when all the 5 arms will be in rest position, no signal will be made.

Hence, the required number of signals $= 1{,}024 - 1 = 1{,}023$.

Example 55: *Find the number of way in which 3 friends can stay in 2 hotels ?*

Solution: Two friends can stay in the same hotel but same friend cannot stay in two hotel at a time. Hence, hotel is repeatable (R) and friend is non-repeatable (NR).

$$\text{Number of ways} = [R]^{NR} = [2]^3 = 8.$$

Example 56: *Find the number of ways in which Dr. Bansal can give 2 hotels to his 3 sons ?*

Solution: A son can get 2 hotels but a hotel cannot be given again and again.

Hence, son is repeatable (R) but hotel is non-repeatable (NR).

$$\text{Number of ways} = [R]^{NR} = (3)^2 = 9.$$

Example 57: *Find the number of ways in which 6 pictures can be hung on 9 picture nails (in a row) ?*

Solution: A picture can not hung on two picture nails at a time, but two pictures can hung on one picture nail.

Hence, picture nail is repeatable (R) but picture is non-repeatable (NR).

$$\text{Number of ways} = [R]^{NR} = (9)^6.$$

Example 58: *In how many ways can 8 students be arranged in*

(i) *a line*

(ii) *a circle.*

Solution:

(i) The number of ways in which 8 students can be arranged in a line = $\lfloor\underline{8}$

$$= 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40{,}320.$$

(ii) The number of ways in which 8 students can be arranged in a circle = $\lfloor\underline{8-1} = 7$

$$= 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5{,}040.$$

Example 59: *In how many ways can a party of 4 men and 4 women be seated at a circular table so that no two women are adjacent ?*

Solution: The 4 men can be seated at the circular table such that there is a vacant seat between every pair of men.

The number of ways in which these 4 men can be seated at the circular table = $\lfloor\underline{3} = 6$.

Now, the 4 vacant seats may be occupied by 4 women in

$${}^4P_4 = \lfloor\underline{4} = 24 \text{ ways.}$$

∴ The required number of ways $= 6 \times 24 = 144$.

Example 60: *In how many ways can 7 persons sit around a table so that all shall not have the same neighbours in any arrangements.*

Solution: 7 persons set at a round table in $\lfloor\underline{6}$ ways.

But, in clockwise and anticlockwise arrangements, each person will have the same neighbours.

So, the required number of ways $= \frac{1}{2} \times \lfloor\underline{6} = 360$.

Exercise

1. In how many ways can 5 person occupy 3 vacant seats ?
2. In how many ways can 5 children stand in a queue.
3. In how many ways can 4 different books, one each in Chemistry, Physics, Biology and Mathematics be arranged in a shelf ?
4. Seven students are contesting the election for the presidentship of the students union. In how many ways can their names be listed on the ballot papers ?
5. Find the number of different 4 letter words that can be formed from the letters of the word 'NUMBERS'.

6. Ten students are participating in a race. In how many ways can the first three prizes be won ?

7. In how many ways can 6 boys and 5 girls be arranged for a group photograph if the girls are to sit on chairs in a row and the boys are stand in a row behind them ?

   [**Hint:** 5 girls can sit on chairs in a row in ${}^5P_5 = \lfloor\underline{5} = 120$ ways

   Also, 6 boys can stand behind them in a row in ${}^6P_6 = \lfloor\underline{6} = 720$ ways.

   $\therefore$ The total number of ways $= 120 \times 720 = 86400$.

8. How many words can be formed from the lettes of the word 'SUNDAY' ?

   How many of these begin with D ?

9. In how many ways can the letters of the word 'FAILURE' ?

10. Find the number of permutations of the letters of the word 'ENGLISH'. How many of these begin with E and end with I ?

11. In how many arrangements of word 'GOLDEN' will the vowels never occur together ?

12. In how many ways can 6 books on Economics, 5 on History and 3 on Hindi be arranged on a shelf so that the books on each subject are always together ?

13. A child has plastic toys bearing the digits 2, 2 and 5. How many three digit numbers can it make, using them ?

14. There are three blue balls, four red balls and five green balls. In how many ways can they be arranged in a row ?

15. How many arrangement can be made out of the letters of the word.

    (a) INDIA (b) ENGINEERING (c) EXAMINATION

16. How many numbers can be formed with the help of the digits 5, 4, 3, 5, 3 ?

17. In how many ways can 6 persons be arranged in

    (a) a line (b) a circle

18. There are 5 men and 5 ladies to dine at a round table. In how many ways can they seat themselves so that no two ladies are together ?

Answers

1.	60	**2.**	120	**3.**	24	**4.**	5,040	**5.**	840	**6.**	720
8.	720; 120	**9.**	576	**10.**	5, 040; 120	**11.**	480	**12.**	31,10,400	**13.**	3
14.	2,520	**15.**	(a) 60 (b) 2,77,200 (c) 49,89,600								
16.	30	**17.**	(a) $\lfloor\underline{6}$ (b) 120			**18.**	$\lfloor\underline{4} \times \lfloor\underline{5} = 2{,}880$				

6.3 Combinations

Combinations: Each of the different groups or selections which can be formed by taking some or all of a number of objects, irrespective of their arrangements, is called a combination. The total number of combinations of n distinct objects taking r $(1 \le r \le n)$ at a time is denoted by $C(n, r)$ or ${}^nC_r$ or $\binom{n}{r}$

where, ${}^nC_r$ is defined only when n and r integers such that $n \ge r$ and $n > 0, r \ge 0$.

Example: *The combinations of 4 objects a, b, c, d taking 2 at a time are*

ab, ac, ad, bc, bd, cd

6.4 Difference between a Permutation and a Combination

In combination, only a group is made and the order in which the objects are arranged is immaterial. In permutation, not only a group is formed, but also an arrangement in definite order is considered.

Note: We used the word 'arrangement' for permutation and selections for combinations.

Theorem 9: The number of all combinations of n distinct objects, taken r at a time, is given by

$$ {}^nC_r = \frac{\lfloor\underline{n}}{\lfloor\underline{r}\ \lfloor\underline{n-r}} $$

Proof: Let the number of all combinations of n objects, taken r at a time, be x.

Then, $\quad {}^nC_r = x.$

Now, each combination contains r objects, which may be arranged amongst themselves is $\lfloor\underline{r}$ ways.

Thus, each combination gives rise to $\lfloor\underline{r}$ permutations.

$\therefore$ x combinations will give rise to $x \times \lfloor\underline{r}$ permutations.

So, the number of permutations of n things, taken r at a time is $x \times \lfloor\underline{r}$

Hence, $\quad {}^nP_r = x \times \lfloor\underline{r} = {}^nC_r \times \lfloor\underline{r}$

$$ \therefore \quad {}^nC_r = \frac{{}^nP_r}{\lfloor\underline{r}} = \frac{\lfloor\underline{n}}{\lfloor\underline{r}\ \lfloor\underline{n-r}} \qquad \left[\because {}^nP_r = \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}}\right] $$

Note: We can write $\quad {}^nC_r = \dfrac{n(n-1)(n-2) \ldots . n \text{ factors}}{\lfloor\underline{r}}.$

Theorem 10: Let $0 \le r \le n$. Prove that ${}^nC_r = {}^nC_{n-r}$.

Proof: We have $${}^nC_{n-r} = \frac{\lfloor\underline{n}}{\lfloor\underline{n-r}\;\lfloor\underline{n-(n-r)}} = \frac{\lfloor\underline{n}}{\lfloor\underline{r}\;\lfloor\underline{n-r}} = {}^nC_r.$$

Theorem 11: To prove that

$${}^nC_r + {}^nC_{r-1} = {}^{(n+1)}C_r.$$

Proof: We have,

$${}^nC_r + {}^nC_{r-1} = \frac{\lfloor\underline{n}}{\lfloor\underline{r}\;\lfloor\underline{n-r}} + \frac{\lfloor\underline{n}}{\lfloor\underline{r-1}\;\lfloor\underline{n-(r-1)}}$$

$$= \frac{\lfloor\underline{n}}{\lfloor\underline{r}\;\lfloor\underline{n-r}} + \frac{\lfloor\underline{n}}{\lfloor\underline{r-1}\;\lfloor\underline{n-r+1}}$$

$$= \frac{\lfloor\underline{n}\cdot(n-r+1)}{\lfloor\underline{r}\cdot\lfloor\underline{(n-r+1)}} + \frac{\lfloor\underline{n}\cdot r}{\lfloor\underline{r}\;\lfloor\underline{n-r+1}}$$

$$= \left\{\frac{\lfloor\underline{n}}{\lfloor\underline{r}\;\lfloor\underline{n-r+1}}\right\}\cdot\{n-r+1+r\}$$

$$= \frac{(n+1)\;\lfloor\underline{n}}{\lfloor\underline{r}\;\lfloor\underline{n-r+1}} = \frac{\lfloor\underline{n+1}}{\lfloor\underline{r}\;\lfloor\underline{n+1-r}} = {}^{n+1}C_r$$

Theorem 12: If $1 \le r \le n$, prove that

$$n \times {}^{(n-1)}C_{r-1} = (n-r+1) \times {}^nC_{r-1}.$$

Proof: We have

$$n \times {}^{(n-1)}C_{r-1} = n \times \frac{\lfloor\underline{n-1}}{\lfloor\underline{r-1} \times \lfloor\underline{(n-1)-(r-1)}}$$

$$= \frac{n \times \lfloor\underline{n-1}}{\lfloor\underline{r-1}\;\lfloor\underline{n-r}}$$

$$= \frac{\lfloor\underline{n}\,(n-r+1)}{\lfloor\underline{r-1} \times \lfloor\underline{n-r} \times (n-r+1)}$$

$$= (n-r+1) \times \frac{\lfloor\underline{n}}{\lfloor\underline{r-1} \times \lfloor\underline{n-r+1}}$$

$$= (n-r+1) \times {}^nC_{r-1}$$

Theorem 13: If n and r are positive integers such that $1 \le r \le n$ then

$$\frac{{}^nC_r}{{}^nC_{r-1}} = \frac{n-r+1}{r}.$$

Proof: We know that

$${}^nC_r = \frac{\underline{\lfloor n}}{\underline{\lfloor r}\,\underline{\lfloor n-r}} \quad \text{and} \quad {}^nC_{r-1} = \frac{\underline{\lfloor n}}{\underline{\lfloor r-1} \times \underline{\lfloor n-r+1}}$$

$$\therefore \qquad \frac{{}^nC_r}{{}^nC_{r-1}} = \frac{\underline{\lfloor n}}{\underline{\lfloor r}\,\underline{\lfloor n-r}} \times \frac{\underline{\lfloor r-1}\,\underline{\lfloor n-r+1}}{\underline{\lfloor n}}$$

$$= \frac{\underline{\lfloor r-1} \times (n-r+1)\,\underline{\lfloor n-r}}{r \times \underline{\lfloor (r-1)}\,\underline{\lfloor n-r}}$$

$$= \left(\frac{n-r+1}{r}\right)$$

Theorem 14: If $1 \le r \le n$, prove that

$${}^nC_r + {}^nC_{r+1} = {}^{n+1}C_{r+1}.$$

Proof:

$${}^nC_r + {}^nC_{r+1} = \frac{\underline{\lfloor n}}{\underline{\lfloor r}\,\underline{\lfloor n-r}} + \frac{\underline{\lfloor n}}{\underline{\lfloor (r+1)}\,\underline{\lfloor n-r-1}}$$

$$= \frac{\underline{\lfloor n}}{\underline{\lfloor r} \times (n-r)\,\underline{\lfloor n-r-1}} + \frac{\underline{\lfloor n}}{(r+1)\,\underline{\lfloor r} \times \underline{\lfloor n-r-1}}$$

$$= \frac{\underline{\lfloor n}}{\underline{\lfloor r}\,\underline{\lfloor n-r-1}}\left(\frac{1}{n-r} + \frac{1}{r+1}\right)$$

$$= \frac{\underline{\lfloor n}}{\underline{\lfloor r}\,\underline{\lfloor n-r-1}} \times \frac{(n+1)}{(n-r)\,(r+1)} = \frac{(n+1)\,\underline{\lfloor n}}{(r+1)\,\underline{\lfloor r} \times (n-r) \times \underline{\lfloor n-r-1}}$$

$$= \frac{\underline{\lfloor n+1}}{\underline{\lfloor r+1}\,\underline{\lfloor n-r}} = \frac{\underline{\lfloor n+1}}{\underline{\lfloor r+1}\,\underline{\lfloor (n+1)-(r+1)}} = {}^{n+1}C_{r+1}$$

Theorem 15: Prove that ${}^nC_p = {}^nC_q \Rightarrow p = q$ or $p + q = n$.

Proof: We have,

$${}^nC_p = {}^nC_q = {}^nC_{n-q}$$

$$\Rightarrow \qquad p = q \quad \text{or} \quad p = n - q \quad \text{or} \quad p + q = n.$$

Evaluate

Example 61: *(i)* ${}^{100}C_{97}$, *(ii)* ${}^{60}C_{60}$, *(iii)* ${}^{5}C_{3}$.

Solution:

(i) $${}^{100}C_{97} = {}^{100}C_{100-97} = {}^{100}C_{3} = \frac{100 \times 99 \times 98}{3 \times 2 \times 1} = 1{,}61{,}700$$

(ii) $${}^{60}C_{60} = 1$$

(iii) $${}^{5}C_{3} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$$

Example 62: *Find n, if*

(i) ${}^{n}C_{7} = {}^{n}C_{9}$ *(ii)* ${}^{n}C_{n-4} = 15$

(iii) ${}^{24}C_{n+5} = {}^{24}C_{3n-1}$ *(iv)* ${}^{2n}C_{2} : {}^{n}C_{3} = 9 : 2.$

Solution:

(i) ${}^{n}C_{7} = {}^{n}C_{9}$ and $7 \neq 9 \Rightarrow n = 7 + 9 = 16.$

(ii) $${}^{n}C_{n-4} = 15 \Rightarrow \frac{\lfloor\underline{n}}{\lfloor\underline{n-4}\ \lfloor\underline{n-(n-4)}} = 15$$

$$\Rightarrow \frac{\lfloor\underline{n}}{\lfloor\underline{n-4}\ \lfloor\underline{4}} = 15$$

$$\Rightarrow \lfloor\underline{n} = 15 \times \lfloor\underline{4}\ \lfloor\underline{n-4}$$

$$\Rightarrow n(n-1)(n-2)(n-3) = 360$$

$$\Rightarrow n = 6.$$

(iii) $${}^{24}C_{n+5} = {}^{24}C_{3n-1}$$

$$\Rightarrow n + 5 = 3n - 1 \text{ or } 24 = n + 5 + (3n - 1)$$

$$\Rightarrow n = 3 \text{ or } n = 5\ .$$

(iv) We have,

$${}^{2n}C_{2} : {}^{n}C_{3} = 9 : 2$$

$$\frac{\lfloor\underline{2n}}{\lfloor\underline{2}\ \lfloor\underline{2n-2}} : \frac{\lfloor\underline{n}}{\lfloor\underline{3}\ \lfloor\underline{n-3}} = \frac{9}{2}$$

$$\Rightarrow \frac{2n(2n-1)}{2} : \frac{n(n-1)(n-2)}{\lfloor\underline{3}} = \frac{9}{2}$$

$$\Rightarrow \qquad \frac{6\,(2n-1)}{(n-1)\,(n-2)} = \frac{9}{2}$$

$$\Rightarrow \qquad 8n - 4 = 3n^2 - 9n + 6 \text{ or } 3n^2 - 17n + 10 = 0$$

$$\Rightarrow \qquad (n-5)\,(3n-2) = 0 \text{ or } n = 5, n = \frac{2}{3}.$$

But only $n = 5$ satisfies the given equation.

Hence, $n = 5$.

Example 63: *If* $^nP_r = 720$ *and* $^nC_r = 120$, *find r.*

Solution: We know that $^nC_r = \dfrac{^nP_r}{\lfloor\underline{r}}$

$$\therefore \qquad 120 = \frac{720}{\lfloor\underline{r}} \quad \Rightarrow \quad \lfloor\underline{r} = \frac{720}{120} = \lfloor\underline{3}$$

Hence, $r = 3$.

Example 64: *Prove that* $^{2n}C_n = \dfrac{2^n \times [1\cdot 3\cdot 5\ \ldots.\ (2n-1)]}{\lfloor\underline{n}}$.

Solution: $^{2n}C_n = \dfrac{\lfloor\underline{2n}}{\lfloor\underline{n}\,\lfloor\underline{2n-n}} = \dfrac{\lfloor\underline{2n}}{(\lfloor\underline{n})^2}$

$$= \frac{(2n)\,(2n-1)\,(2n-2)\ \ldots\ldots\ 4\cdot 3\cdot 2\cdot 1}{(\lfloor\underline{n})^2}$$

$$= \frac{[2n\,(2n-2)\,(2n-4)\ \ldots.\ 4\cdot 2] \times [(2n-1)\,(2n-3)\ \ldots.\ 5\cdot 3\cdot 1]}{(\lfloor\underline{n})^2}$$

$$= \frac{2^n\,[n\,(n-1)\,(n-2)\ \ldots.\ 2\cdot 1]\,[(2n-1)\,(2n-3)\ \ldots.\ 5\cdot 3\cdot 1]}{(\lfloor\underline{n})^2}$$

$$= \frac{2^n\,\lfloor\underline{n}\,[1\cdot 3\cdot 5\ \ldots.\ (2n-3)\,(2n-1)]}{(\lfloor\underline{n})^2}$$

$$= \frac{2^n \times [1\cdot 3\cdot 5\ \ldots.\ (2n-3)\,(2n-1)]}{\lfloor\underline{n}}.$$

Example 65: *If* $^nC_{r-1} : {^nC_r} : {^nC_{r+1}} :: 3 : 4 : 5$ *find n and r.*

Solution: $\dfrac{^nC_{r-1}}{^nC_r} = \dfrac{3}{4}$ and $\dfrac{^nC_r}{^nC_{r+1}} = \dfrac{4}{5}$

$$\Rightarrow \qquad \frac{r}{n-r+1} = \frac{3}{4} \text{ and } \frac{r+1}{n-r} = \frac{4}{5}$$

$$\Rightarrow \qquad 3n - 7r = -3 \text{ and } 4n - 9r = 5$$

$$\Rightarrow \qquad r = 27, n = 62$$

Example 66: If ${}^{107}C_{87} = {}^{107}C_x$*, then find the value of x.* **[B.C.A. (Meerut) 2007]**

Solution: Given

$$ {}^{107}C_{87} = {}^{107}C_x \quad \text{...(1)} $$

We know that

$$ {}^{n}C_r = {}^{n}C_{n-r} \quad \text{...(2)} $$

Comparing (1) and (2), we get

$$ x = n - r = 107 - 87 = 20 $$

Thus, $x = 20$

Example 67: If ${}^{1270}C_{1117} + {}^{1270}C_{1116} = {}^{x}C_{1117}$*, then find the value of* x.

Solution: Given

$$ {}^{1270}C_{1117} + {}^{1270}C_{1116} = {}^{x}C_{1117} \quad \text{...(1)} $$

We know that.

$$ {}^{n}C_r + {}^{n}C_{r-1} = {}^{(n+1)}C_r \quad \text{...(2)} $$

Comparing (1) and (2), we get

$$ x = n + 1 \Rightarrow x = 1270 + 1 = 1271 $$

Thus, $x = 1271$

Example 68: *If* $x\,.\,{}^{77}C_{55} = 23 \times {}^{78}C_{55}$*, then find the value of* x.

[B.C.A. (Bundelkhand) 2010]

Solution: Given

$$ x\,.\,{}^{77}C_{55} = 23 \times {}^{78}C_{55} \quad \text{...(1)} $$

Equation (1) can be re-written as:

$$ x\,.\,{}^{77}C_{55} = (78 - 56 + 1)\,{}^{78}C_{55} \quad \text{...(2)} $$

We know that,

$$ n\,.\,{}^{(n-1)}C_{(r-1)} = (n - r + 1)\,{}^{n}C_{(r-1)} \quad \text{...(3)} $$

Comparing (2) and (3), we get

$$ x = n = 78 $$

Hence, $x = 78$

Example 69: If ${}^{15}C_{11} + {}^{15}C_x = {}^{16}C_{12}$ *Find x.* **[B.C.A. (Avadh) 2008]**

Solution: We have

$$ {}^{15}C_{11} + {}^{15}C_x = {}^{16}C_{12} \qquad \text{...(1)}$$

Equation (1) can be re-written as:

$$ {}^{15}C_{11} + {}^{15}C_x = {}^{(15+1)}C_{(11+1)} \qquad \text{...(2)}$$

We know that

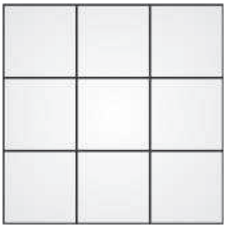
$$ {}^{n}C_r + {}^{n}C_{r+1} = {}^{(n+1)}C_{(r+1)} \qquad \text{...(3)}$$

Comparing (2) and (3), we get

$$x = r + 1 = 11 + 1 = 12$$

Hence, $x = 12$

Example 70:



In how many different ways can all of 5 identical balls be placed in the cells shown above such that each row contains at least 1 ball ? **[B.C.A. (Purvanchal) 2010]**

Solution: The following combinations for the given rows can be made:

(I) 3 balls in first row can be put in ${}^3C_3$ ways

1 ball in second row can be in ${}^3C_1$ ways

1 ball in third row can be put in ${}^3C_1$ ways.

∴ In this case the total number of ways the balls can be put in the squares

$$= {}^3C_3 \times {}^3C_1 \times {}^3C_1$$

$$= 1 \times 3 \times 3 = 9$$

Similarly, we have

(II) ${}^3C_2 \times {}^3C_2 \times {}^3C_1 = 3 \times 3 \times 3 = 27$

(III) ${}^3C_1 \times {}^3C_3 \times {}^3C_1 = 3 \times 1 \times 3 = 9$

(IV) ${}^3C_1 \times {}^3C_1 \times {}^3C_3 = 3 \times 3 \times 1 = 9$

(V) ${}^3C_2 \times {}^3C_1 \times {}^3C_2 = 3 \times 3 \times 3 = 27$

(VI) ${}^3C_1 \times {}^3C_2 \times {}^3C_2 = 3 \times 3 \times 3 = 27$

∴ Total number of arrangements

$$= 9 + 27 + 9 + 9 + 27 + 27$$

$$= 108$$

Example 71: *In how many ways can a committee of 5 members be selected from 6 men and 5 ladies, consisting of 3 men and 2 ladies ?*

Solution: 3 men out of 6 and 2 ladies out of 5 can be selected in

$$^{6}C_3 \times {}^{5}C_2 = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} \times \frac{5 \times 4}{2 \times 1} = 200 \text{ ways.}$$

Example 72: *A committee of 5 is to be formed out of 6 men and 4 ladies. In how many ways can this be done, when*

(a) atleast 2 ladies are included

(b) atmost 2 ladies are included.

Solution:

(a) We have to make selection of

(i) (2 ladies out of 4) and (3 men out of 6) or

(ii) (3 ladies out of 4) and (3 men out of 6) or

(iii) (4 ladies out of 4) and (1 man out of 6)

The number of ways of these selections are

Case (i) $^{4}C_2 \times {}^{6}C_3 = 6 \times 20 = 120$

Case (ii) $^{4}C_3 \times {}^{6}C_2 = 4 \times 15 = 60$

Case (iii) $^{4}C_4 \times {}^{6}C_1 = 1 \times 6 = 6.$

Total number of ways to select such a committee = 120 + 60 + 6 = 186

(b) We have to make a selection of

(i) (1 lady out of 4) and (4 men out of 6) or

(ii) (2 ladies out of 4) and (3 men out of 6)

The number of ways of these selections are

Case (i) $^{4}C_1 \times {}^{6}C_4 = 4 \times 15 = 60$

Case (ii) $^{4}C_2 \times {}^{6}C_3 = 6 \times 20 = 120$

Case (iii) number of lady is selected then

Number of ways of selecting committee = $6 c_5 = 6$

∴ the required number of ways = 60 + 120 = 180.

Example 73: *An examination paper containing* 12 *questions consists of two parts, A and B, part A contains* 7 *questions and part B contains* 5 *questions. A candidate is required to attempt* 8 *questions, selecting atleast* 3 *from each part. In how many ways the candidate select the questions ?*

Solution: We have,

(i) (3 out of 7 from A) and (5 out of 5 from B) or

(ii) (4 out of 7 from A) and (4 out of 5 from B) or

(iii) (5 out of 7 from A) and (3 out of 5 from B)

The number of ways of these selections are

Case (i) $^7C_3 \times {}^3C_5 = 35 \times 1 = 35$

Case (ii) $^7C_4 \times {}^5C_4 = 35 \times 5 = 175$

Case (iii) $^7C_5 \times {}^5C_3 = 21 \times 10 = 210$

$\therefore$ the number of ways of selecting the questions $= (35 + 175 + 210) = 420$.

Example 74: *If* 20 *lines are drawn in a plane such that no two of them are parallel and no three are concurrent, then in how many points will they intersect each other ?*

Solution: Since, no two lines are parallel and no three pass through the same point, their point of intersection, (*i.e.*, number of ways of selecting two lines from 20 lines)

$$= {}^{20}C_2 = \frac{20 \times 19}{2 \times 1} = 190.$$

Example 75: *How many lines can be drawn through* 21 *points on a circle ?*

Solution: One line can be drawn by joining any 2 points, from 21 given points.

$\therefore$ Total no. of lines $= {}^{21}C_2 = \dfrac{21 \times 20}{2 \times 1} = 210.$

Example 76: *There are* 15 *points in a plane, no three of which are collinear. Find the number of triangles formed by joining them.*

Solution: One triangle is formed by selecting a group of 3 points from 15 given points.

$\therefore$ No. of triangles $= {}^{15}C_3 = \dfrac{15 \times 14 \times 13}{3 \times 2 \times 1} = 455.$

Example 77: *Everybody in a room shakes hands with everybody else. The total number of handshakes is 66. How many people are there in the room ?*

Solution: Let there are n persons in the room.

$\therefore$ Total no. of handshakes is same as the number of ways of selecting 2 persons among n persons, *i.e.*, $^nC_2$.

$$\therefore \qquad {}^nC_2 = 66 \quad \text{(given)}$$

$$\frac{n(n-1)}{2} = 66 \Rightarrow n^2 - n - 132 = 0$$

$$\Rightarrow \qquad (n-12)(n+11) = 0 \Rightarrow n = 12 \text{ or } n = -11$$

But, no. of persons cannot be negative

$$n = 12.$$

Example 78: *Determine the number of 5 cards combination out of a deck of 52 cards, if there is exactly one ace in each combination.*

Solution: One ace can be selected from 4 aces and 4 other cards can be selected from remaining 48 cards.

$$\therefore \qquad \text{No. of ways} = {}^4C_1 \times {}^{48}C_4$$

$$= 4 \times \frac{48 \times 47 \times 46 \times 45}{4 \times 3 \times 2 \times 1}$$

$$= 7{,}78{,}320.$$

Example 79: *From a class of 25 students, 10 are to be chosen for an excursion party. There are 3 students who decide that either all of them will join or none of them will join. In how many ways can they be chosen ?*

Solution: There are two possibilities:

Case I: When the 3 particular students join the party.

In this case, we have to select 7 more students from the remaining 22 students. This can be done in $^{22}C_7$ **ways**.

Case II: When the 3 particular students do not join the party.

In this case, we have to select 10 students from the remaining 22 students.

This can be done in $^{22}C_{10}$ ways

$$\text{So, total number of ways} = {}^{22}C_7 + {}^{22}C_{10}$$

$$= 170544 + 646646$$

$$= \frac{22 \times 19 \times 17 \times 3 \times 8}{10 \times 9 \times 8}(720 + 2{,}730) = 8{,}17{,}190.$$

Example 80: *A committee of 12 is to be formed from 9 women and 8 men. In how many ways this can be done, if atleast 5 women have to be included in a committee ? In how many of these committees :*

(a) the women are in majority, (b) the men are in majority ?

Solution: The committee can be selected in the following way:

9 Women	5	6	7	8	9	⟶	(at least 5 women)
8 Men	7	6	5	4	3		

$$\therefore \quad \text{Total, number of ways} = {}^9C_5 \times {}^8C_7 + {}^9C_6 \times {}^8C_6 + {}^9C_7 \times {}^8C_5 + {}^9C_8 \times {}^8C_4 + {}^9C_9 \times {}^8C_3$$

$$= 1{,}008 + 2{,}352 + 2{,}016 + 630 + 56$$

$$= 6{,}062$$

Number of committees in which women are in majority

$$= 2{,}016 + 630 + 56 = 2{,}702$$

Number of committees in which men are in majority = 1,008.

Example 81: *If m parallel lines in a plane are intersected by a family of n parallel lines, find the number of parallelograms formed.*

Solution: A parallelogram is formed when two lines from the set of *m* parallel lines are intersected by the other 2 lines from another set of *n* parallel lines.

Hence, the number of parallelograms formed in this case

$$= {}^mC_2 \times {}^nC_2$$

$$= \frac{m(m-1)}{2} \times \frac{n(n-1)}{2}$$

$$= \frac{mn(m-1)(n-1)}{4}.$$

Example 82: *A student is allowed to select at most n books from a collection of (2n + 1) books. If the total number of ways in which he can select at least one book is 255, find the value of n.*

Solution: The student may select at most *n* books out of (2*n* + 1) books. Thus, according to the given condition,

$${}^{2n+1}C_1 + {}^{2n+1}C_2 + {}^{2n+1}C_3 + \ldots. + {}^{2n+1}C_n = 255$$

$$\Rightarrow \quad \frac{1}{2}[2 \cdot {}^{2n+1}C_1 + 2 \cdot {}^{2n+1}C_2 + 2 \cdot {}^{2n+1}C_3 + \ldots. + 2 \cdot {}^{2n+1}C_n] = 255$$

$$\Rightarrow \quad \frac{1}{2}[({}^{2n+1}C_1 + {}^{2n+1}C_1) + ({}^{2n+1}C_2 + {}^{2n+1}C_2) + \ldots. + ({}^{2n+1}C_n + {}^{2n+1}C_n)] = 255$$

$$\Rightarrow \quad \frac{1}{2}[({}^{2n+1}C_1 + {}^{2n+1}C_{2n}) + ({}^{2n+1}C_2 + {}^{2n+1}C_{2n-1}) + \ldots. + ({}^{2n+1}C_n + {}^{2n+1}C_{n+1})] = 255 \quad (\because {}^mC_r = {}^mC_{m-r})$$

$$\Rightarrow \quad \frac{1}{2}[({}^{2n+1}C_1 + {}^{2n+1}C_2 + {}^{2n+1}C_3 + \ldots. + {}^{2n+1}C_{2n})] = 255$$

$$\Rightarrow \quad \frac{1}{2}[({}^{2n+1}C_0 + {}^{2n+1}C_1 + {}^{2n+1}C_2 + \ldots + {}^{2n+1}C_{2n+1}) - ({}^{2n+1}C_0 + {}^{2n+1}C_{2n+1})] = 255$$

$$\Rightarrow \quad \frac{1}{2}[2^{2n+1} - (1+1)] = 255 \quad (\because {}^nC_0 + {}^nC_1 + {}^nC_2 + \ldots. + {}^nC_n = 2^n)$$

$\Rightarrow \qquad 2^{2n+1} - 2 = 510$

$\Rightarrow \qquad 2^{2n+1} = 512$

$\Rightarrow \qquad 2^{2n+1} = 2^9$

$\Rightarrow \qquad 2n + 1 = 9$

$\Rightarrow \qquad n = 4.$

Example 83: *A committee of 5 is to be formed out of 6 gents and ladies. In how many ways this can be done when*

(i) at least 2 ladies are included,

(ii) at most 2 ladies are included.

Solution: Total gents = 6, Total ladies = 4

1. At least 2 ladies out of 4 ladies are included in the following ways
   (i) 3 gents out of 6 gents + 2 ladies out of 4 ladies = ${}^6C_3 \times {}^4C_2$ ways
   (ii) 2 gents out of 6 gents + 3 ladies out of 4 ladies = ${}^6C_2 \times {}^4C_3$
   (iii) 1 gent out of 6 gents + 4 ladies out of 4 ladies = ${}^6C_1 \times {}^4C_4$

   Thus, the total number of ways of forming the committee

$$= {}^6C_3 \times {}^4C_2 + {}^6C_2 \times {}^4C_3 + {}^6C_1 \times {}^4C_4 = 120 + 60 + 6 = 186$$

2. At most 2 ladies out of 4 ladies are included in the following ways:
   (i) No lady and 5 gents = ${}^6C_5$ ways
   (ii) 1 lady and 4 gents = ${}^4C_1 \times {}^6C_4$ ways
   (iii) 2 ladies and 3 gents = ${}^4C_2 \times {}^6C_3$ ways

   Thus, the required ways of forming committee

$$= {}^6C_5 + {}^4C_1 \times {}^6C_4 + {}^4C_2 \times {}^6C_3$$

$$= 6 + 60 + 120 = 186.$$

Example 84: *A polygon has 44 diagonals. Find the numbers of its sides.*

Solution: The number of diagonals of n sided polygon

$$= {}^nC_2 - n = \frac{n(n-3)}{2}$$

Given that: $\qquad \dfrac{n(n-3)}{2} = 44$

$\Rightarrow \qquad n^2 - 3n = 88$

$\Rightarrow \qquad n^2 - 3n - 88 = 0$

$\Rightarrow \qquad (n-11)(n+8) = 0$

$\Rightarrow \qquad n = 11 \qquad (\because n > 0)$

Example 85: *A box contains 5 different red and 6 different white balls. In how many ways can 6 balls be selected so that there are at least two balls of each colour ?*

Solution: Total red balls $= 5$

Total white balls $= 6$

6 balls can be selected from the box so that there are at least two balls of each colour in the following ways:

(i) 2 red balls and 4 white balls $= {}^5C_2 \times {}^6C_4$ ways

(ii) 3 red balls and 3 white balls $= {}^5C_3 \times {}^6C_3$ ways

(iii) 4 red balls and 2 white balls $= {}^5C_4 \times {}^6C_2$

Hence, the required ways $= {}^5C_2 \times {}^6C_4 + {}^5C_3 \times {}^6C_3 + {}^5C_4 \times {}^6C_2$

$$= 150 + 200 + 75 = 425$$

Example 86: *In how many ways can* 21 *identical books on English and* 19 *identical books on Hindi be placed in a row on a shelf so that two books on Hindi may not be together ?*

Solution: Let E denote the position of an English book and H of Hindi book on a shelf. Since no two books of Hindi must be together on the shelf therefore, we first arrange all books of English in a row. Since all the books of English are identical so they can be arranged in a row in only one way as shown below:

$$H\ E\ H\ E\ H\ E\ H \ldots\ldots H\ E\ H\ E$$

As there are 21 books of English so there will be 22 places for Hindi books.

Now we have 19 identical books of Hindi, therefore out of 22 places, 19 places.

$$\text{Total no. of ways} = {}^{22}C_{19} = \frac{\lfloor 22}{\lfloor 19 \lfloor 22-19} = \frac{22 \times 21 \times 20}{3 \times 2 \times 1} = 1540$$

Example 87: *In how many ways can a cricket team be selected from a group of* 25 *players containing* 10 *batsmen, 8 bowlers, 5 all-rounders and 2 wicket keepers ? Assume that the team of* 11 *players requires 5 batsman, 3 all-rounders, 2 bowlers and 1 wicket keeper.*

Solution: The selection of team is divided into 4 phases:

(i) 5 batsman out of 10 batsman can be selected in ${}^{10}C_5$ ways.

(ii) 3 all-rounders out of 5 all-rounders can be selected in ${}^5C_3$ ways.

(iii) 2 bowlers out of 8 bowlers can be selected in ${}^8C_2$ ways.

(iv) 1 wicket keeper out of 2 wicket keepers can be selected in ${}^2C_1$ ways.

$\therefore$ Team can be selected in ${}^{10}C_5 \times {}^5C_3 \times {}^8C_2 \times {}^2C_1$ ways $= 1{,}41{,}120$ ways.

Example 88: *In how many ways can a committee of 3 ladies and 4 gentlemen be appointed from a group consisting of 8 ladies and 7 gentlemen ? What will be the number of ways, if Aishwarya of this committee refuses to serve in a committee in which Salman is a member ?*

Solution:

(i) 3 ladies can be selected from 8 ladies in ${}^8C_3$ ways and 4 gentlemen from 7 gentlemen in ${}^7C_4$ ways.

∴ No. of ways in which the committee can be appointed

$$= {}^8C_3 \times {}^7C_4 = 56 \times 35 = 1{,}960.$$

(ii) First we find the no. of committees of 3 ladies and 4 gentlemen in which both Aishwarya and Salman are members. In this case, we can select 2 other ladies from the remaining 7 in ${}^7C_2$ ways and 3 other gentlemen from remaining 6 in ${}^6C_3$ ways.

∴ No. of ways in which both Aishwarya and Salman are always included

$$= {}^7C_2 \times {}^6C_3 = 21 \times 20 = 420.$$

Hence, the required no. of committees in which Aishwarya and Salman do not serve together

$$= 1{,}960 - 420 = 1{,}540.$$

Example 89: *In how many ways a committee of 5 persons can be formed out of 7 Indians and 4 Pakistanis when the committee contains:*

(i) All Indians

(ii) At least two Pakistanis

(iii) At least two Indians

(iv) Not more than two Pakistanis

(v) Not more than two Indians **[B.C.A. (Meerut) 2008]**

Solution:

(i) When committee consists of all its members as Indians. The number of ways selecting the committee = ${}^7C_5 \times {}^4C_0 = 21$

(ii) When at least two Pakistanis are selected, then the committee can be selected as follows:

Case I: When 2 Pakistanis out of 4 and 3 Indians out of 7 are selected. Then number of ways selecting the committee = ${}^4C_2 \times {}^7C_3 = 210$

Case II: Number of ways selecting 3 Pakistanis and 2 Indians = ${}^4C_3 \times {}^7C_2 = 84$

Case III: Number of ways selecting 4 Pakistanis and 1 Indian = ${}^4C_4 \times {}^7C_1 = 7$

∴ Total number of ways selecting the committee

$$= 210 + 84 + 7 = 301$$

(iii) When at least two Indians are selected. Then number of ways selecting the committee

$$= {}^7C_2 \times {}^4C_3 + {}^7C_3 \times {}^4C_2 + {}^7C_4 \times {}^4C_1 + {}^7C_5 \times {}^4C_0$$

$$= 84 + 210 + 168 + 21 = 483$$

(iv) When not more than two Pakistanis are selected. Then number of ways selecting the committee $= {}^4C_1 \times {}^7C_4 + {}^4C_2 \times {}^7C_3$

$$= 140 + 840 = 980$$

(v) When not more than two Indians are selected.

Then number of ways selecting the committee

$$= {}^7C_1 \times {}^4C_4 + {}^7C_2 \times {}^4C_3 = 7 + 84 = 91$$

Exercise

1. Verify that ${}^8C_4 + {}^8C_3 = {}^9C_4$.

2. (i) If ${}^nC_7 = {}^nC_5$, find n.

   (ii) If ${}^{20}C_r = {}^{20}C_{r+6}$, find r.

   (iii) If ${}^nP_r = 1680$ and ${}^nC_r = 70$, find n and r.

   (iv) If ${}^{n+1}C_{r+1} : {}^nC_r = 11 : 6$ and ${}^nC_r : {}^{n-1}C_{r-1} = 6 : 3$, find n and r.

3. If ${}^{16}C_r : {}^{15}C_{r-1} = 10 : 5$ find r.

4. If ${}^{18}C_r = {}^{18}C_{r+2}$, find ${}^rC_5$.

5. Find ${}^{25}C_{22} - {}^{24}C_{21}$.

6. In how many ways can a cricket team be chosen out of a batch of 15 players, if

   (a) there is no restriction on the selection

   (b) a particular player is always chosen

   (c) a particular player is never chosen.

   [**Hint:** (a) ${}^{15}C_{11}$ (b) ${}^{14}C_{10}$ (c) ${}^{14}C_{11}$]

7. Out of 5 men and two women, a committee of 3 is to be formed. In how many ways can it be formed if at least one women is to be included ?

8. From a class of 32 students, 4 are to be chosen for a competition. In how many ways can this be done ?

9. In how many ways can a student choose 5 courses out of 9 courses if 2 courses are compulsory for every student ?

10. A question paper has two parts, part 'A' and part 'B', each containing 10 questions. If the student has to choose 8 from part 'A' and 5 from part 'B'. In how many ways can he choose the questions ?

11. From a class of 25 students, 10 are to be chosen for an excursion party. There are 3 students who decide that either all of them will join or none of them will join. In how many ways can the ten be chosen ?

12. A cricket team of 11 players is to be selected from 16 players including 5 bowlers and 2 wicket keepers. In how many ways can a team be selected so as to consist of exactly 3 bowlers and 1 wicket keeper ?

13. (i) How many straight lines can be obtained by joining 12 points, 5 of which are collinear ?

(ii) How many triangles can be obtained by joining 12 points, five of which are collinear ?

14. A polygon has 44 diagonals. Find the number of its sides.

15. In a village, there are 87 families of which 52 families have at most 2 children. In a rural development programme 20 families are to be chosen for assistance, of which at least 18 families must have at most 2 children. In how many ways can the choice be made ?

16. A bag contains 4 red, 3 white and 2 blue marbles. 3 marbles are drawn at random. Determine the number of ways of selecting at least 1 white marble in this selection.

17. In how many ways can 5 white balls and 4 black balls be arranged in a row so that no two black balls are together.

18. A man has 6 friends. In how many ways can he invite one or more to a party ?

19. A boy has 3 library tickets and 8 books of his interest in the library of these 8, he does not want to borrow chemistry part II, unless chemistry part I is also borrowed.

20. Out of 6 teachers and 8 students, a committee of 11 is to be formed. In how many ways can this be done, if the committee contains

(i) exactly 4 teachers,

(ii) at least 4 teachers.

21. In how many ways can a cricket team be selected from 17 players in which 5 players can bowl ? Each cricket team must include 2 bowlers ?

Answers

2.	(i) 12 (ii) 7 (iii) $n = 8, r = 4$ (iv) $n = 10, r = 5$					**3.**	8
4.	56	**5.**	276	**6.**	(a) 1,365 (b) 1,001 (c) 364	**7.**	25
8.	35,960	**9.**	35	**10.**	11,340	**11.**	1,464
12.	722	**13.**	(i) 57 (ii) 210	**14.**	${}^{52}C_{18} \times {}^{35}C_2 + {}^{52}C_{19} \times {}^{35}C_1 + {}^{52}C_{20}$	**16.**	6417.20
18.	63	**19.**	41	**20.**	(i) 120 (ii) 344	**21.**	2,200

❑❑❑

Chapter 7

Probability

7.1 Introduction

There are two approaches to the theory of probability viz. (a) Classical approach and (b) Set-theoretic or Axiomatic approach. We begin the chapter with the classical approach.

7.1.1 Classical Approach to Probability

(A) **Random Experiment and Events**: A random experiment may be defined as an experiment which when repeated under essentially identical conditions does not give unique results but may result in any one of the several possible outcomes. These outcomes are known as events or cases. Events are denoted as A, B, C etc.

For Example:

(*i*) Getting a head (H) or a tail (T) is an event when we toss a coin.

(*ii*) Getting any of the six faces 1, 2, 3, 4, 5, 6 is an event we throw a die.

(*iii*) Getting an ace or a king or a queen is an event when we draw a card from a pack of well-shuffled cards.

(B) **Exhaustive Events**: The total number of possible outcomes in a random experiment are known as exhaustive events.

For Example:

(*i*) There are 2 exhaustive events viz. head (H) and tail (T) when we toss a coin.

(*ii*) There are 6 exhaustive events when we throw a die.

(*iii*) When we throw two dice the exhaustive number of events is $6^2 = 36$ given as follow:

(1, 1) , (1, 2) , (1, 3) , (1, 4) , (1, 5) , (1, 6) ;
(2, 1) , (2, 2) , (2, 3) , (2, 4) , (2, 5) , (2, 6) ;
(3, 1) , (3, 2) , (3, 3) , (3, 4) , (3, 5) , (3, 6) ;
(4, 1) , (4, 2) , (4, 3) , (4, 4) , (4, 5) , (4, 6) ;
(5, 1) , (5, 2) , (5, 3) , (5, 4) , (5, 5) , (5, 6) ;
(6, 1) , (6, 2) , (6, 3) , (6, 4) , (6, 5) , (6, 6) ;

(*iv*) When we draw two cards from a pack of playing cards, the exhaustive number of events is ${}^{52}C_2$, since 2 cards can be drawn out of 52 cards in ${}^{52}C_2$ ways.

(C) Favourable Events: The events which cause the happening of a particular event A are called the favourable events to the event A.

For Example:

(*i*) There are three favourable events for the occurrence of an even number (or an odd number) in the throwing of a die.

(*ii*) When we draw a card from a pack of cards, there are 4 favourable events for the drawing of an ace ; there are 12 favourable events for the drawing of a face card (king, queen, jack).

(*iii*) When we throw two dice together, the number of events favourable to getting a sum of 9 is 4 given by the following combinations :

(3, 6), (4, 5), (6, 3), (5, 4)

(D) Mutually Exclusive Events: Such events where the occurrence of one rules out the occurrence of the other, are called mutually exclusive events.

For Example:

(*i*) In tossing a coin there are two mutually exclusive events, for if head comes in a trial, then tail cannot come in the same trial or vice versa.

(*ii*) In throwing a die there are 6 mutually exclusive events, for if any of the six faces comes, then the possibility of other faces in the same trial is ruled out.

(*iii*) There are 52 mutually exclusive events in drawing a card from a pack of cards.

(E) Equally Likely Events: The events are said to be equally likely if none of them is expected to occur in preference to other.

For Example:

(*i*) In tossing a coin, there are 2 equally likely viz. H and T.

(*ii*) In tossing a die, there are 6 equally likely events viz. 1, 2, 3, 4, 5, 6.

7.1.2 Classical Definition of Probability

If there are n exhaustive, mutually exclusive and equally likely events out of which m are favourable to the happening of an event A, then the probability of the happening of A, denoted by $P[A]$ is defined as

$$P[A] = \frac{\text{Favourable number of cases}}{\text{Exhaustive number of cases}} = \frac{m}{n}$$

Remarks:

1. It is clear that if m cases out of n exhaustive cases favour the happening of an event A, then $n - m$ cases favour that the event A will not happen. Thus, the probability that the event A will not happen denoted by $P[\overline{A}]$, is given by

$$P[\overline{A}] = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P[A]$$

$$\therefore \quad P[A] + P[\overline{A}] = 1 \qquad \text{...(1)}$$

[B.C.A. (Meerut) 2003]

2. Since $0 \le m \le n$, so $0 \le \frac{m}{n} \le 1$

Hence, $0 \le P[A] \le 1$, for any event A.

If $P[\overline{A}] = 0$, then A is called an impossible or null event.

If $P[A] = 1$, then A is called a certain or sure event.

3. The probability $P[A]$ of the happening of an event A is also known as the **probability of success**, denoted by p and the probability $P[\overline{A}]$ of the non-happening of the event A is known as the **probability of failure** denoted by q.

From (1), it follows that:

$$p + q = 1, (0 \le p \le 1,\ 0 \le q \le 1).$$

4. If A and B are two events, then the probability of the happening of A or B (*i.e.* at least one of the two events) is denoted as $P[A + B]$ or $P[A \cup B]$. The probability of the simultaneous occurrence of two events A and B is denoted by $P[AB]$ or $P[A \cap B]$.

5. The following results from permutations and combinations are very useful:

   (i) $n! = n(n-1) \ldots 3, 2, 1, n! = n(n-1)! = n(n-1)(n-2)!$ etc.

   (ii) The number of **permutations** of n different things taken r at a time is

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

   (*iii*) The number of **combinations** of n different things taken r at a time is $^{n}C_{r}$ or $\binom{n}{r}$ where

$$^{n}C_{r} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

For Example:

$$\binom{6}{2} = {}^{6}C_2 = \frac{6\,!}{2\,!4\,!} = \frac{6\,.\,5}{2\,.\,1} = 15, \quad \binom{10}{3} = {}^{10}C_3 = \frac{10\,.\,9\,.\,8}{3\,.\,2\,.\,1} = 120$$

$$\binom{52}{4} = {}^{52}C_4 = \frac{52\,.\,51\,.\,50\,.\,49}{4\,.\,3\,.\,2\,.\,1} = 270725$$

Solved Examples

Example 1: *What is the chance that a leap year selected at random will contain 53 sundays?*

Solution: A leap year consisting of 366 days has 52 complete weeks and 2 days extra. The following are the possible combinations for there two extra days:

(i) Sunday and Monday

(ii) Monday and Tuesday

(iii) Tuesday and Wednesday

(iv) Wednesday and Thursday

(v) Thursday and Friday

(vi) Friday and Saturday and

(vii) Saturday and Sunday

In order that a leap year selected at random should contain 53 Sundays, one of the two extra days must be Sunday. Since out of the above 7 possibilities 2 viz; (*i*) and (v*ii*) are favourable to this event, the required probability $= \frac{2}{7}$

Example 2: *(i) Two cards are drawn at random from a well-shuffled pack of 52 cards. Find the probability of getting 2 aces.*

Solution: Total number of cases in which 2 cards can be drawn out of 52 cards is $n = {}^{52}C_2$. There are 4 aces in a pack of 52 cards and out of 4 aces, 2 aces can be drawn is $m = {}^{4}C_2$ ways.

$$\text{Required Probability} = \frac{m}{n} = \frac{{}^{4}C_2}{{}^{52}C_2} = \frac{4\,.\,3}{52\,.\,51} = \frac{1}{221}$$

(ii) In a single throw of '2' dice, what is probability of obtaining a total 10 ?

[B.C.A. (Meerut) 2003]

Solution: Total numbers of cases $= 6 \times 6 = 36$

Find problem of obtaining a total of 10.

Favourable outcome $= (4, 6), (5, 5), (6, 4)$

Total number of favourable cases = 3.

$\therefore$ The probability of throwing $10 = \frac{3}{36} = \frac{1}{12}$

Example 3: *A bag contain 6 white, 7 red and 5 black balls. Find the chance that three balls drawn at random are all white.*

Solution:

$$\text{Total number of cases} = {}^{18}C_3 = \frac{18 \times 17 \times 16}{3 \times 2 \times 1} = 816$$

$$\text{Favourable number of cases} = {}^{6}C_3 = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

$$\text{Required probability} = \frac{20}{816} = \frac{5}{204}$$

Example 4: *Find the probability of getting an even number in a throw of a single die.*

Solution: A die has six faces bearing numbers 1 to 6.

Total number of exhaustive cases $= 6$ *i.e.* $n = 6$

out of these six numbers, there are three even numbers viz. 2, 4 and 6.

Total number favourable cases $= 3$ *i.e.* $m = 3$.

$$\text{Required probability} = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}$$

Example 5: *From a pack of 52 cards, two cards are drawn at random. Find the chance that one is a king and the other a queen.*

Solution: Total number of cases is $n = {}^{52}C_2$

Since there are 4 kings and 4 queens, so the number of favourable cases is $m = {}^{4}C_1 \times {}^{4}C_1$.

$$\text{Required probability} = \frac{m}{n} = \frac{{}^{4}C_1 \times {}^{4}C_1}{52_{C_2}} = \frac{4.4}{52.51}.2.1 = \frac{8}{663}$$

Example 6: *Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. Obtain the chance that at the most 2 of them are children.*

[B.C.A. (Meerut) 2002]

Solution: We have a group of 3 men, 2 women and 4 children and we have to choose 4 persons in a way. Four persons can be chosen in the following ways:

(i) 1 child, 2 man, 1 woman probability $= \frac{{}^{4}C_1 \times {}^{3}C_2 \times {}^{2}C_1}{{}^{9}C_4}$

(ii) 1 child, 1man, 2 women probability $= \frac{{}^{4}C_1 \times {}^{3}C_1 \times {}^{2}C_2}{{}^{9}C_4}$

(iii) 2 children, 1 man, 1 woman probability $= \dfrac{{}^4C_2 \times {}^3C_1 \times {}^2C_1}{{}^9C_4}$

(iv) 0 child, 3 men, 1 woman probability $= \dfrac{{}^4C_0 \times {}^3C_3 \times {}^2C_1}{{}^9C_4}$

(v) 0 child, 2 men, 2 women probability $= \dfrac{{}^4C_0 \times {}^3C_2 \times {}^2C_2}{{}^9C_4}$

Since only one case is possible so required probability will be the sum of all these cases $= \dfrac{77}{126}$

Example 7: *A ball is drawn at random from a box containing 6 red balls, 4 white balls and 5 blue balls. Determine the probability that the ball drawn is:*

(i) red *(ii) white*

(iii) blue *(iv) not red*

(v) red or white. **[B.C.A. (Meerut) 2009]**

Solution: Let R, W and B denote the events of drawing a red ball, a white ball and a blue ball respectively.

(i) $P(R) = \dfrac{\text{Number of favourable outcomes}}{\text{Total number of outcomes}} = \dfrac{n(R)}{n(S)}$

$$= \frac{6}{6+4+5} = \frac{6}{15} = \frac{2}{5} = 0.4$$

(ii) $P(W) = \dfrac{6}{6+4+5} = \dfrac{4}{15} = 0.26$

(iii) $P(B) = \dfrac{5}{6+4+5} = \dfrac{5}{15} = \dfrac{1}{3} = 0.33$

(iv) $P(\overline{R}) = 1 - P(R) = 1 - \dfrac{2}{5} = \dfrac{3}{5} = 0.6$

(v) $P(R+W) = P(R) + P(W) = \dfrac{2}{5} + \dfrac{4}{15} = \dfrac{10}{15} = \dfrac{2}{3} = 0.66$

Example 8: *If four ladies and six gentlemen sit for a photograph in a row at random what is the probability that no two ladies will sit together ?*

Solution: 4 ladies + 6 gentlemen = 10 persons is all

They can sit in a row in ${}^{10}P_{10} = 10!$ ways

Then, $n(S) = 10!$

Now, if no two ladies are to be together, the ladies have 7 position, 2 at ends and 5 between the gentlemen arrangement $L, G_1, L, G_2, L, G_3, L, G_4, L, G_5, L, G_6, L$. The 4 ladies can occupy these 7 seats in ${}^7P_4$ ways and 6 gentlemen can occupy 6 seats in ${}^6P_6$ ways.

Therefore, the total number of these arrangements $= {}^7P_4 \times {}^6P_4$ ways.

$$\text{Therefore, the required probability} = \frac{\text{No. of favourable outcomes}}{\text{Total no. of outcomes}}$$

$$= \frac{{}^7P_4 \times {}^6P_4}{10\,!} = \frac{1}{6} \text{ or } 0.17$$

Example 9: *In a class there are* 13 *students* 5 *of them are boys and the rest are girls. Find the probability that two students selected at random will be both girls.*

Solution: Two students out of 13 can be selected in ${}^{13}C_2$ ways and two girls out of 8 can be selected in ${}^8C_2$ ways.

$$\text{Therefore, required probability} = \frac{{}^8C_2}{{}^{13}C_2} = \frac{14}{39} = 0.36$$

Example 10: *A box contains* 5 *white balls,* 4 *black balls and* 3 *red balls. Three balls are drawn randomly. What is the probability that they will be (i) white, (ii) black, (iii) red ?*

Solution: Let W, B and R denote the events of drawing three white, three black and three red balls respectively.

(i) $$P(W) = \frac{n(W)}{n(S)} = \frac{{}^5C_3}{(5+4+3)\,C_3} = \frac{{}^5C_3}{{}^{12}C_3} = \frac{1}{22} = 0.045$$

(ii) $$P(B) = \frac{n(B)}{n(S)} = \frac{{}^4C_3}{{}^{12}C_3} = \frac{1}{55} = 0.018$$

(iii) $$P(R) = \frac{n(R)}{n(S)} = \frac{{}^3C_3}{{}^{12}C_3} = \frac{1}{1320} = 0.76 \times 10^{-4}$$

Example 11: *A bag contains* 6 *red,* 5 *white and* 4 *black balls. Two balls are drawn. Find the probability that none of them is red.* **[B.C.A. (Rohilkhand) 2006]**

Solution: Total number of balls $= 6 + 5 + 4 = 15$

$$\text{Number of exhaustive events} = {}^{15}C_2 = \frac{15 \times 14}{2 \times 1} = 105$$

Total number of non red balls $= 5 + 4 = 9$

$$\therefore \quad \text{Number of favourable events} = {}^9C_2$$

$$= \frac{9 \times 8}{2 \times 1} = 36$$

$$\therefore \quad \text{Required probability} = \frac{36}{105} = \frac{12}{35}$$

Example 12: *If from a lottery of* 30 *tickets, marked* 1 *to* 30 *four tickets be drawn, what is the probability that those marked* 1 *and* 2 *are among them ?* **[B.C.A. (Meerut) 2010, 2009]**

Solution: Four tickets can be drawn out of 30 in $30\,C_4$ ways. If the ticket marked 1 and 2 are to be always included, then 2 more can be drawn out of the remaining 28 in $28\,C_2$ ways.

$$\therefore \qquad \text{Required probability} = \frac{28\,C_2}{30\,C_4}$$

$$= \frac{\dfrac{28\times27}{2\times1}}{\dfrac{30\times29\times28\times27}{4\times3\times2\times1}}$$

$$= \frac{28\times27\times4\times3\times2\times1}{30\times29\times28\times27\times2\times1} = \frac{2}{145}$$

Example 13: *A drawer contains* 50 *balls and* 150 *nuts. Half of the bolts and half of the nuts are rusted. If one item is chosen at random, what is the probability that it is rusted or in a bolt ?* **[B.C.A. (Garhwal) 2010]**

Solution: Let ‘A’ = {the item is rusted}

and ‘B’ = {the item is a bolt}

Here, A and B are non-mutually exclusive events

$$\therefore \qquad P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

$$= \frac{100}{200} + \frac{50}{200} - \frac{25}{200} = \frac{150-25}{200} = \frac{125}{200} = \frac{5}{8}$$

$$\therefore \qquad \text{Required probability} = \frac{5}{8}$$

Example 14: *In a single throw of two dice, find the probability that neither a doublet (same number on both the dice) nor a total of* 9 *will appear.* **[B.C.A. (Lucknow) 2008, 2004]**

Solution: Number exhaustive cases $= 6\times6 = 36$ cases favorable to a doublet are:

$(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)$ *i.e.* 6 ways

$\therefore$ P (probability of getting a doublet) $= \frac{6}{36} = \frac{1}{6}$ cases favorable to a total of 9 cases are

$(3,6), (6,3), (4,5), (5,4)$ *i.e.* 4 ways

$\therefore$ P (probability of getting a total 9) $= \frac{4}{36} = \frac{1}{9}$

Since the two events are mutually exclusive

P (a doublet or a total of 9) = P (a doublet) + P (a total of 9)

$$= \frac{1}{6} + \frac{1}{9} = \frac{3+2}{18} = \frac{5}{18}$$

$\therefore$ P(neither of doublet nor a total of 9) $=1-P$ (a doublet nor a total of 9)

$$=1-\frac{5}{18}=\frac{18-5}{18}=\frac{13}{18}.$$

Example 15: *A class consists of 80 students, 25 of them are girls and 55 boys, 10 of them are rich and the remaining poor, 20 of them are fair complexioned. What is the probability of selecting a fair complexioned rich girl ?* **[B.C.A. (Lucknow) 2006; B.C.A. (Kanpur) 2005]**

Solution: Probability of selecting a fair complexioned person

$$=\frac{20}{80}=\frac{1}{4}$$

Probability of selecting a rich person

$$=\frac{10}{80}=\frac{1}{8}$$

Probability of selecting a girl $=\frac{25}{80}=\frac{5}{16}$

Since these events are independent of each other, therefore the probability of selecting a fair complexioned rich girl $=\frac{1}{4}\times\frac{1}{8}\times\frac{5}{16}=\frac{5}{512}=0.009$

Example 16: *A bag contains 6 white and 9 black balls. Four balls are drawn at a time. Find the probability for the first draw to give four white and second draw to give four black balls in each of the following cases:*

(i) The balls are replaced before the second draw

(ii) The balls are not replaced before the second draw

Solution:

(i) When the balls are replaced before the second draw. Here two draws are independent

$\therefore$ $P(A\cap B)=P(A)\times P(B)$

Probability of drawing 4 white balls $=P(A)=\frac{6\,C_4}{15\,C_4}$

Probability of drawing 4 black balls $=P(B)=\frac{9\,C_4}{15\,C_4}$

Now, $P(A\cap B)=P(A)\times P(B)=\frac{6\,C_4}{15\,C_4}\times\frac{9\,C_4}{15\,C_4}$

$$=\frac{6\times5\times4\times3}{15\times14\times13\times12}\times\frac{9\times8\times7\times6}{15\times14\times13\times12}=\frac{6}{5915}$$

(ii) When balls are not replaced before the second draw.

Here two events are not independent.

Probability of drawing 4 white balls $= \frac{9\,C_4}{15\,C_4} = \frac{1}{91}$

Now, the balls drawn have not been replaced, therefore the bag contains 2 white and 9 black balls.

Probability of drawing 4 black balls

$$P\,(B|A) = \frac{9\,C_4}{11C_4} = \frac{21}{55}$$

$$\therefore \qquad P(A \cap B) = P(A)\,P(B|A) = \frac{6\,C_4}{15\,C_4} \times \frac{9\,C_4}{11C_4} = \frac{1}{91} \times \frac{21}{55} = \frac{3}{715}$$

Example 17: *An win contains 7 red and 4 blue balls. Two balls are drawn at random with replacement. Find the probability of getting bag.* **[B.C.A. (A.M.U.) 2011]**

(i) 2 red balls (ii) 2 blue balls (iii) are red and one blue ball.

[B.C.A. (Rohtak) 2010]

Solution:

(i) P (2 red balls) $= \frac{7}{11} \times \frac{7}{11} = \frac{49}{121}$

(ii) P (2 blue balls) $= \frac{4}{11} \times \frac{4}{11} = \frac{16}{121}$

(iii) P (one red and one blue ball) $= \frac{7}{11} \times \frac{4}{11} + \frac{4}{11} \times \frac{7}{11}$

$$= \frac{28}{121} + \frac{28}{121} = \frac{56}{121}$$

Example 18: *Two dice are thrown. Find the probability that a multiple of 2 occurs on one dice and multiple of 3 occurs on the other.* **[B.C.A. (Agra) 2007]**

Solution: Two dice can be thrown in $6 \times 6 = 36$ days

Multiples of 2 are 2, 4, 6 and multiple of 3 are 3, 6.

Favourable cases for getting multiple of 2 on one dice and multiple of 3 on the other are:

$(2,3), (2,6), (4,3), (4,6), (6,3), (6,6), (6,4), (6,2), (3,2), (3,4)$ and $(3,6)$ *i.e.* 11 ways.

$\therefore$ required probability $= \frac{11}{36}$.

Example 19: *In a single throw of three dice, find the probability of getting a total of 17 or 18.* **[B.C.A. (Meerut) 2008]**

Solution: Three dice can be thrown in $6 \times 6 \times 6 = 216$ ways 4 total of 17 can be obtained as:

$$(5,6,6), (6,5,6), (6,6,5)$$

A total of 18 can be obtained as (6,6,6)

$\therefore$ No. of class favorable to a total of 17 or 18 are 4

$\therefore$ P (a total of 17 or 18) $= \frac{4}{216} = \frac{1}{54}$

Example 20: *Two dice are thrown. Getting two numbers whose sum is divisible by* 4 *or* 5 *is considered a success. Find the probability of success.* **[B.C.A. (Agra) 2005]**

Solution: To obtain sum divisible by 4 we must get a total of 4 or 8 or 12 and to obtain sum divisible by 5, we obtain a total of 5 or 10. Hence the probability of success is P(4 or 8 or 12 or 5 or 20) $= P(4) + P(8) + P(12) + P(5) + P(10)$.

A total of 4 can be obtained as $(1,3), (3,1), (2,2)$ *i.e.* 3 ways

$$\therefore \qquad P(4) = \frac{3}{36}$$

Similarly, $\quad P(8) = \frac{5}{36}, P(12) = \frac{1}{36}, P(5) = \frac{4}{36}$ and $P(10) = \frac{3}{36}$

Hence the required probability $= \frac{3}{36} + \frac{5}{36} + \frac{1}{36} + \frac{4}{36} + \frac{3}{36}$

$$= \frac{16}{36} = \frac{4}{9}.$$

Example 21: *A bag contains tickets numbered from* 1 *to* 20. *Two tickets are drawn. Find the probability that both the numbers are prime.* **[B.C.A. (Kurukshetra) 2011, 2008]**

Solution: Total numbers of cards = 20

The probability of getting 2 cards from 20 $= {}^{20}C_2$

The prime number in between 1 to 20 = 8

The probability of getting both number to be prime $= {}^{8}C_2$

$\therefore$ required probability $= \frac{{}^{8}C_2}{{}^{20}C_2} = \frac{8 \times 7}{20 \times 19} = \frac{28}{190}$

Example 22: *Two cards are drawn from a pack of* 52 *cards, one after another without replacement. Find the chance that one of these is an ace and other is queen of opposite shade.*

Solution: There are two possibilities:

(i) The first card drawn is an ace and the second drawn is a queen of opposite shade (colour). The probability of this compound event $= \frac{4}{52} \times \frac{2}{51}$

(ii) The first card is a queen and the second is an ace of opposite shade. The probability of this event $= \frac{4}{52} \times \frac{2}{51}$.

Hence, the required probability.

$$= \frac{4}{52} \times \frac{2}{51} + \frac{4}{52} \times \frac{2}{51} = \frac{2}{13 \times 51} + \frac{2}{13 \times 51}$$

$$= \frac{2}{663} + \frac{2}{663} = \frac{4}{663}$$

Example 23: Tim Tim and Vishu appear for an interview for two vacancies. The probability of Tim Tim's selection is $\frac{1}{3}$ and that of Vishu's selection is $\frac{1}{5}$. Find the probability that only one of them will be elected.

Solution: P (Tim Tim selection) $= \frac{1}{3}$

$$P(\text{Tim Tim selection}) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P\ (\text{Vishu's selection}) = \frac{1}{5}$$

$$P\ (\text{Vishu's not selection}) = 1 - \frac{1}{5} = \frac{4}{5}$$

$$\therefore \qquad \text{required probability} = \frac{1}{3} \times \frac{4}{5} + \frac{1}{5} \times \frac{2}{3}$$

$$= \frac{4}{15} + \frac{2}{15} = \frac{6}{15} = \frac{2}{5}$$

7.1.3 Theorem of Total Probability

If n events $A_1, A_2, \ldots, A_n$ are not mutually exclusive, then the probability of the happening of at least one of the events is the sum of the probabilities of the individual events. In symbols

$$P\,[A_1 + A_2 + \ldots + A_n] = p\,[A_1] + P\,[A_2] + \ldots + P\,[A_n]$$

Or

$$P\,[A_1 \cup A_2 \cup \ldots \cup A_n] = P\,[A_1] + P\,[A_2] + \ldots + P\,[A_n]$$

Proof: Let N be the total number of mutually exclusive, exhaustive and equally likely out of which m_1, cases are favourable to A_1, m_2 cases are favourable to A_2 and so on.

The probability of occurrence of the event $A_1 = P\,[A_1] = \frac{m_1}{N}$

The probability of occurrence of the event $A_2 = P\,[A_2] = \frac{m_2}{N}$ ...(1)

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

The probability of occurrence of the event $A_n = P\,[A_n] = \frac{m_n}{N}$

The events being mutually exclusive and equally likely, the total number of cases favourable to the event A_1 or A_2 or ... A_n is

$$m = m_1 + m_2 + \ldots + m_n \qquad \text{...(2)}$$

$$\therefore \qquad P\,[A_1 + A_2 \ldots + A_n] = \frac{m}{N}$$

$$= \frac{m_1 + m_2 + \ldots m_n}{N} \quad \text{using} \qquad \text{...(3)}$$

$$= \frac{m_1}{N} + \frac{m_2}{N} \ldots + \frac{m_n}{N}$$

$$= P\,[A_1] + P\,[A_2] + \ldots P\,[A_n]$$

7.1.4 Compound Events

The simultaneous occurrence of two or more events is called a **compound event.**

Conditional Probability

The probability of the happening of an event A when the event B has already happened is called the **conditional probability** and is denoted by $P\left[\frac{A}{B}\right]$.

Similarly, $P\left[\frac{B}{A}\right]$ means the probability of the happening of an event B when the event A has already happened.

Independent Events

Two events are said to be **independent** if the probability of the happening of one does not depend on the happening or non-happening of the other.

7.1.5 Theorem of Compound Probability or Multiplicative Law of Probability

If A and B are two events, then

$$P\,[AB] = P\,[A]\,P\left[\frac{B}{A}\right]$$

Proof: Suppose there are n exhaustive, mutually exclusive and equally likely cases, out of which m cases are favourable to the happening of the event A.

$$\therefore \qquad P\,[A] = \frac{m}{n} \qquad ...(1)$$

Let m_1 out comes be favourable to the happening of the event B, Then

$$P\left[\frac{B}{A}\right] = \frac{m_1}{m} \qquad ...(2)$$

Now out of n exhaustive, mutually exclusive and equally likely out comes, m_1 are favourable to the happening of A and B

The probability of the simultaneous occurrence of A and B is

$$P\,[AB] = \frac{m_1}{n} = \frac{m_1}{m} \times \frac{m}{n} = \frac{m}{n} \times \frac{m_1}{m}$$

$$\text{Hence,} \qquad P\,[AB] = P\,[A]\,P\left[\frac{B}{A}\right], \text{ using (1) and (2)}$$

Cor. 1: Interchanging A and B in the above result, we get

$$P\,[BA] = P\,[B]\,P\left[\frac{A}{B}\right]$$

Hence, $$P[AB] = P[B]\,P\left[\frac{A}{B}\right] \qquad [\because P[AB] = P[BA]]$$

Cor. 2: If A and B are independent events, then

$$P[AB] = P[A]\,P[B]$$

Proof: We have $P\left[\frac{B}{A}\right] = P[B]$, since A and B are independent. Hence

$$P[AB] = P[A]\,P[B]$$

Cor. 3: If $A_1, A_2, \ldots, A_n$ are n independent events, then

$$P[A_1 A_2 \ldots A_n] = P[A_1]\,P[A_2] \ldots P[A_n]$$

Cor. 4: If $p_1, p_2, \ldots, p_n$ are the probabilities that certain events happen, then the probabilities of their non-happening are $1 - p_1, 1 - p_2 \ldots, 1 - p_n$. The probability of all the events not happening $(1 - p_1)(1 - p_2) \ldots (1 - p_n)$.

Hence, the probability that at least one of these events happens

$$= 1 - [(1 - P_1)(1 - P_{2)} \ldots (1 - P_n)]$$

Example 24: *What is the chance of throwing a total of* 5 *or* 11 *with two dice ?*

Solution: Exhaustive number of cases $= 6^2 = 36$

Let $P[A]$ denote the probability of getting a total of 5 and $P[B]$ the probability of getting a total of 11.

Favourable cases of getting a total of 5 are

(1, 4); (4, 1); (2 , 3); (3, 2)

$$\therefore \qquad P[A] = \frac{4}{36}$$

Favourable cases of getting a total of 11 are (5, 6) : (6, 5)

$$\therefore \qquad P[B] = \frac{2}{36}$$

The probability of getting a total of 5 or 11 are

$$P[A + B] = P[A] + P[B] = \frac{4}{36} + \frac{2}{36} = \frac{6}{36} = \frac{1}{6}$$

Example 25 : *It is* 8 : 5 *against a husband who is* 55 *years old living till he is* 75 *and* 4 : 3 *against his wife who is now* 48, *living till she is* 68 *Find the probability that*

(i) *the couple will be alive for* 20 *years hence*

(ii) *one at least of these will be alive for* 20 *years since.*

Solution: Let A and B respectively denote the events of the husband and wife living 20 years hence. Then

$$P[A] = \frac{5}{8+5} = \frac{5}{13}, \quad P[B] = \frac{3}{4+3} = \frac{3}{7}$$

$$\Rightarrow \qquad P[\overline{A}] = \frac{8}{13} \text{ and } P[\overline{B}] = \frac{4}{7}$$

(i) Probability that the couple lives for 20 years

$$= P[AB] = P[A]\,P[B] = \frac{5}{13} \times \frac{3}{7} = \frac{15}{91}$$

(ii) Probability that none of A and B lives $= P[\overline{A}]\,P[\overline{B}] = \frac{8}{13} \times \frac{4}{7}$.

Hence, the probability that at least one of A and B lives

$$= 1 - \frac{8}{13} \times \frac{4}{7} = \frac{59}{91}$$

Example 26: *A problem in statistics is given to the three students A, B and C whose chances of solving it are* $\frac{1}{2}, \frac{1}{3}$ *and* $\frac{1}{4}$ *respectively. What is the probability that the problem is solved?*

Solution: The probabilities that the problem is solved by the students A, B and C are respectively

$$P[A] = \frac{1}{2}, P[B] = \frac{1}{3} \text{ and } P[C] = \frac{1}{4}$$

Now,

$$P[\overline{A}] = \frac{1}{2},\ P[\overline{B}] = \frac{2}{3},\ P[\overline{C}] = \frac{3}{4}$$

The problem will be solved if at least one of then solves it. The probability that none of them solves it ,is

$$= P[\overline{A}]\,P[\overline{B}]\,P[\overline{C}] = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4} = \frac{1}{4}$$

Hence, the probability that the problem is solved

$$= 1 - \frac{1}{4} = \frac{3}{4}$$

Similar Example: *A problem in statistics in given to the three students A, B and C whose chances of solving it are* $\frac{1}{2}, \frac{3}{4}$ *and* $\frac{1}{4}$ *respectively. What is the probability that the problem will be solved ?*

$\left(\frac{29}{32}\right)$. **Ans.**

Example 27: *p is the probability that a man aged x will die within a year. Find the probability that out of five men A, B, C, D, E each aged x, A will die during the year and the first to die.*

Solution: Since the probability that a man aged x will die in a year is p, the probability that the man does not die in a year $= (1 - p)$. The probability that none of the five men each aged x dies in a year $= (1 - p)^5$. The probability that at least one of the 5 men will die during the year $= \{1 - (1 - p)^5\}$. It may be noted that one man may be any one of A, B, C, D, E. The probability that A will be the first to die $= \frac{1}{5}$. Hence, the probability that A will die during the year and be the first to die.

$$= \frac{1}{5}\{1 - (1 - p)^5\}$$

Similar Example: *p is the probability that a man aged x will die in a year. Find the probability that out of n men $A_1, A_2, \ldots, A_n$ each aged x, A_1 will die in a year and be the first to die.*

$$\left[\frac{1}{n}\{1-(1-p)^n\}\ \textbf{Ans.}\right]$$

Example 28: *A speaks truth in 75% and B in 80% of the cases. In what percentage of cases are they likely to contradict each other in stating the same fact ?*

Solution: Probabilities that A and B speak the truth are given by

$$P[A]=\frac{75}{100}=\frac{3}{4}, P[B]=\frac{80}{100}=\frac{4}{5}$$

Probabilities that they do not speak the truth are

$$P[\overline{A}]=1-\frac{3}{4}=\frac{1}{4}, P[\overline{B}]=1-\frac{4}{5}=\frac{1}{5}$$

Probabilities that they contradict each other is

$$=P[A]\,P[\overline{B}]+P[B]\,P[\overline{A}]\ =\frac{3}{4}\times\frac{1}{5}+\frac{4}{5}\times\frac{1}{4}=\frac{7}{20}$$

Hence, A and B are likely to contradict each other in $\frac{7}{20}\times 100=35\%$ of the cases.

Example 29: *Probability that a boy will pass an examination is $\frac{3}{5}$ and that for a girl is $\frac{2}{5}$. What is the probability that at least one of them passes ?*

Solution: The probability that the boy passes is $P[A]=\frac{3}{5}$.

The Probability that the girl passes is $P[B]=\frac{2}{5}$.

The probability that at least one of them passes is

$$=P[A+B]=P[A]+P[B]-P[AB]$$

$$=p[A]+P[B]-P[A]\,P[B]$$

$$=\frac{3}{5}+\frac{2}{5}-\frac{3}{5}.\frac{2}{5}=\frac{19}{25}$$

Example 30: *If from a bag of 6 red, 5 white and 4 black balls two balls are drawn, find probability that none of them is white.* **[B.C.A. (Meerut) 2003]**

Solution: Given 6 Red Balls, 5 White Balls, 4 Black Balls

Total number of balls $=6+5+4=15$ balls

2 balls drawn

Find none of them is white.

$$=\frac{^6C_2}{^{15}C_2}+\frac{^4C_2}{^{15}C_2}+\frac{^6C_1\times{}^4C_1}{^{15}C_2}$$

$$= \frac{\left(\frac{6\,!}{2\,!4\,!}\right)}{\left(\frac{15\,!}{2\,!13\,!}\right)} + \frac{\left(\frac{4\,!}{2\,!2\,!}\right)}{\left(\frac{15\,!}{2\,!13\,!}\right)} + \frac{\left(\frac{6\,!}{2\,!}\right)\left(\frac{4\,!}{3\,!}\right)}{\left(\frac{15\,!}{2\,!13\,!}\right)}$$

$$= \left[\frac{\left(\frac{6\times 5}{2}\right)}{\left(\frac{15\times 14}{2}\right)}\right] + \left[\frac{6}{105}\right] + \left[\frac{(6\times 4)}{\frac{15\times 14}{2}}\right]$$

$$= \frac{15}{105} + \frac{6}{105} + \frac{24}{105}$$

$$= \frac{45}{105} = \frac{3}{7}$$

Example 31: *A bag contains 6 white and 9 black balls. Four balls are drawn at a time. Find the probability for the first draw to give 4 white and the second to give 4 black balls in each of the following cases:*

(i) *The balls are not replaced before the second draw.*

(ii) *The balls are replaced before the second draw.* **[B.C.A. (Agra) 2004]**

Solution:

(i) We can draw 4 balls from a bag containing 6 white and 9 black balls in $^{15}C_4$ ways.

Let A be the event that the first draw gives 4 white balls and B be the event that the second draw gives 4 black balls. We draw 4 white balls out of 6 white balls in $^{6}C_4$ ways.

$$\therefore \qquad P\,[A] = \frac{^{6}C_4}{^{15}C_4}$$

Now, if the drawn balls are not replaced, we can draw the other 4 balls in $^{11}C_4$ ways. The event B that the second draw results in 4 black balls (on the assumption that the first draw has given 4 white balls) has $^{9}C_4$ favourable cases.

$$\therefore \qquad P\left[\frac{B}{A}\right] = \frac{^{9}C_4}{^{11}C_4}$$

Hence, $$P\,[AB] = P\,[A] \times P\left[\frac{A}{B}\right] = \frac{^{6}C_4}{^{15}C_4} \times \frac{^{9}C_4}{^{11}C_4} = \frac{3}{715}$$

(ii) We have $$P\,[A] = \frac{^{6}C_4}{^{15}C_4}$$

Since the balls are replaced after the first draw, the probability of drawing 4 black balls in the second draw is $\frac{^{9}C_4}{^{15}C_4}$.

$\therefore$ $P[AB] = P[A].P[B]$ as A and B are independent events.

Hence, $$P[AB] = \frac{^{6}C_4}{^{15}C_4} \times \frac{^{9}C_4}{^{15}C_4} = \frac{3}{2963}$$

Example 32: *A and B toss a coin alternately on the understanding that the first to obtain head wins the toss. Show that their respective chances of winning are* 2/3 *and* 1/3.

Solution: Probabilities that A and B get head in a toss are given by

$$P[A] = P[B] = \frac{1}{2}$$

$$\therefore \quad P[\overline{A}] = [\overline{B}] = 1 - \frac{1}{2} = \frac{1}{2}$$

If A begins, he can win in following mutually exclusive ways

$$(A), (\overline{A}\ \overline{B}\ A), (\overline{A}\ \overline{B}\ \overline{A}\ \overline{B}\ A), \ldots$$

Therefore, by the theorem of total probability, probability that A wins

$$= P[A] + P[\overline{A}\ \overline{B}\ A] + P[\overline{A}\ \overline{B}\ \overline{A}\ \overline{B}\ A] + \ldots$$

$$= P[A] + P[\overline{A}]\,P[\overline{B}]\,P[A] + \{P[\overline{A}]\,P[\overline{B}]\}^2 P[A] + \ldots$$

$$= \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \ldots = \frac{a}{1-r}, \quad \left[a = \frac{1}{2}, r = \left(\frac{1}{2}\right)^2\right]$$

$$= \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$$

Since total probability is unity and one of the players is to win. Probability that B wins $1 - \frac{2}{3} = \frac{1}{3}$.

Example 33: *A husband and wife appear in an interview for two vacancies in the same post. The probability of the husband's selection is* $\frac{1}{7}$ *and that of wife's selection is* $\frac{1}{5}$. *What is probability that only one of them will be selected ?* **[B.C.A. (Meerut) 2009]**

Solution: The probability of the husband's selection is

$$P[H] = \frac{1}{7} \Rightarrow P[\overline{H}] = 1 - \frac{1}{7} = \frac{6}{7}$$

The probability of the wife's selection is

$$P[W] = \frac{1}{5} \Rightarrow P[\overline{W}] = 1 - \frac{1}{5} = \frac{4}{5}$$

The probability that only one of them will be selected

$$= P[H]P[\overline{W}] + P[W]P[\overline{H}]$$

$$= \frac{1}{7}\cdot\frac{4}{5} + \frac{1}{5}\cdot\frac{6}{7} = \frac{10}{35} = \frac{2}{7}$$

Example 34: *A and B take turns in throwing two dice, the first to throw 9 being awarded the price. Show that their chance of winning are in the ratio 9 : 8.*

Solution: The total of 9 can be obtained as follows:

(3, 6), (4, 5), (5, 4) and (6, 3). Total number of cases $= 6 \times 6 = 36$.

No. of favourable cases $= 4$

Then, $$P[A] = P[B] = \frac{4}{36} = \frac{1}{9}$$

Therefore, $$P[\overline{A}] = P[\overline{B}] = 1 - \frac{1}{9} = \frac{8}{9}$$

Since A begins, he can win following mutually exclusive ways:

$$(A), (\overline{A}\ \overline{B}\ A), (\overline{A}\ \overline{B}\ \overline{A}\ \overline{B}\ A), \ldots$$

Therefore, by the theorem of total probability, probability that A wins

$$= P[A] + P[\overline{A}\ \overline{B}\ A] + P[\overline{A}\ \overline{B}\ \overline{A}\ \overline{B}\ A) + \ldots$$

$$= \frac{1}{9}\left\{1 + \left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^4 + \ldots\right\}$$

which is a G.P. with C.R. $= \frac{8}{9}$

$$= \frac{1}{9}\cdot\frac{1}{1 - \frac{64}{81}} = \frac{9}{17}$$

Since total probability is unity and one of two players is to win. Probability that B wins $= 1 - \frac{9}{17} = \frac{8}{17}$. Hence, their chances of winning are in the ratio 9 : 8.

Example 35: *Three groups of children contain respectively 3 girls and 1 boy, 2 girls and 2 boys and 1 girl and 3 boys. One child is selected at random from each group. Show that the chance that the three selected consist of 1 girl and 2 boys is 13/32.*

[B.C.A. (Delhi) 2005; B.C.A. (Meerut) 2002]

Solution: 1 girl and 2 boys can be selected from each group in the following mutually exclusive ways:

S.No.	Ist Group	2nd Group	3rd Group
I	Girl	Boy	Boy
II	Boy	Girl	Boy
III	Boy	Boy	Girl

By the theorem of total probability.

$$\text{Required probability} = P\,[I] + P\,[II] + P\,[III] \qquad ...(1)$$

The probability of selecting a girl from the first group is 3/4, of selecting a boy from the second group is 2/4, and of selecting a boy from the third group is 3/4. Since the three events of selecting children from three groups are independent of each other, therefore

$$P\,[I] = \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}$$

Similarly,

$$P\,[II] = \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{3}{32}$$

$$P\,[III] = \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{1}{32}$$

Substituting in (1), the required probability is

$$= \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32}$$

Example 36: *A and B are two independent witness (i.e. there is no collusion between them) in a case. The probability that A will speak the truth is x and the probability that B will speak the truth is y. A and B agree in a certain statement. Show that the probability that this statement is true, is*

$$= \frac{xy}{1 - x - y + 2xy}$$

Solution: Let A_1 be the event that A and B agree in a statement and A_2 be the event that their statement is correct

Then $$P\,[A_1] = xy + (1 - x)\,(1 - y) = 1 - x - y + 2xy$$

Now $$P\,[A_1\; A_2] = xy$$

By compound probability theorem, we have

$$P\,[A_1\; A_2] = P\,[A_1]\, P\left[\frac{A_2}{A_1}\right]$$

$$\therefore \qquad P\left[\frac{A_2}{A_1}\right] = \frac{P\,[A_1\; A_2]}{P\,[A_1]} = \frac{xy}{1 - x - y + 2xy}$$

Example 37: *A bag contains* 50 *tickets numbered* 1,2, ..., 50. *Five tickets are drawn at random and arranged in ascending order of magnitude :* $(x_1 < x_2 < \ldots < x_{50})$. *What is the probability that* $x_3 = 30$ *?*

Solution: Five tickets out of 50 tickets can be drawn in ${}^{50}C_5$ ways. Since $x_3 = 30$, tickets x_1 and x_2 must be drawn from the set {1, 2, ... , 29} and x_4 and x_5 from the set $\{31, 32, \ldots, 50\}$. The tickets x_1, x_2, x_4 and x_5 can be chosen in ${}^{29}C_2 \times {}^{20}C_2$ ways. Hence

$$P\,[x_3 = 30] = \frac{{}^{29}C_2 \times {}^{20}C_2}{{}^{50}C_5}$$

Example 38: *Find the probability of throwing* 10 *exactly in one throw with three dice.*

[B.C.A. (Meerut) 2004]

Solution: We can throw 10 with three dice in the following ways:

(1, 3, 6), (1, 4, 5), (1, 5, 4), (1, 6, 3),
(2, 2, 6), (2, 3, 5), (2, 4, 4), (2, 5, 3), (2, 6, 2),
(3, 1, 6), (3, 2, 5), (3, 3, 4), (3, 4, 3), (3, 5, 2), (3, 6, 1)
(4, 1, 5), (4, 2, 4), (4, 3, 3), (4, 4, 2), (4, 5, 1),
(5, 1, 4), (5, 2, 3), (5, 3, 2), (5, 4, 1),
(6, 1, 3), (6, 2, 2), (6, 3, 1),

The above favourable case are $m = 27$

Total number of exhaustive cases is $= 6 \times 6 \times 6 = 216$

Hence, the required probability $= \frac{27}{216} = \frac{1}{8}$.

Example 39: *Find the probability of a total of* 9 *when three pair dice are tossed once.*

Solution: We get a total of 9 when three fair dice are tossed once in the following 25 ways:

(1, 2, 6), (1, 3, 5), (1, 4, 4), (1, 5, 3) (1, 6, 2),
(2, 1, 6), (2, 2, 5), (2, 3, 4), (2, 4, 3), (2, 5, 2), (2, 6, 1)
(3, 1, 5), (3, 2, 4), (3, 3, 3), (3, 4, 2), (3, 5, 1),
(4, 1, 4), (4, 2, 3), (4, 3, 2), (4, 4, 1), (5, 1, 3),
(5, 2, 2), (5, 3, 1), (6, 1, 2), (6, 2, 1)

Hence, the required probability $= \frac{25}{216}$.

Example 40: *Compare the probability of at least one* 6 *in* 4 *tosses of a fair die with the probability of at least one double -* 6 *in* 24 *tosses of two fair dice.*

Solution: Probability that no. 6 appears in 4 tosses of a fair die is

$$= \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{6}\right) = \left(\frac{5}{6}\right)^4$$

Hence, the probability of at least one 6 in 4 tosses of a fair die is

$$= 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} = \frac{671}{1296} = 0.52 \text{ (approx)}$$

Probability of a double - 6 *i.e.* the pair (6, 6) when a pair of fair dice is thrown $= \frac{1}{36}$. Thus, the probability that no double - 6 appears when a pair of fair dice is thrown

$$= 1 - \frac{1}{36} = \frac{35}{36}$$

and this probability when there are 24 tosses of two fair dice is $\left(\frac{35}{36}\right)^{24}$. Hence, the probability of at least one double - 6 in 24 tosses of two pair dices

$$= 1 - \left(\frac{35}{36}\right)^{24} = 0.49 \text{ (approx)}$$

The ratio of the two probabilities is 52 : 49.

Example 41: *In an assembly-line operation, $\frac{1}{3}$ of the items being produced are defective. If three items are picked at random and tested. What is the probability ?*

(i) That exactly one of them will be defective ?

(ii) That at least one of them will be defective ?

Solution: Probability of a defective item $= \frac{1}{3}$

Probability of a non-defective item $= 1 - \frac{1}{3} = \frac{2}{3}$.

(i) Probability that exactly one of the three items will be defective

$$= \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} + \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} + \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{12}{27} = \frac{4}{9}$$

(ii) Probability that none of the three items will be defective

$$= \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27}$$

The probability that at least one of the items will be defective

$$= 1 - \frac{8}{27} = \frac{19}{27}$$

Example 42: *A bag contains 10 white and 6 black balls. 4 balls are successively drawn out and not replaced. What is the probability that they are alternately of different colours ?*

[B.C.A. (Meerut) 2002]

Solution: The balls are chosen without replacement. There will be two situations first ball is white or first ball is black.

(I) Beginning with White:

The probability of drawing a white ball $= \frac{10}{16}$

The probability of drawing a black ball then $= \frac{6}{15}$

The probability of drawing a white ball then $= \frac{9}{14}$

The probability of drawing a black ball then $= \frac{5}{13}$

The probability of the compound event $= \frac{10}{16} \times \frac{6}{15} \times \frac{9}{14} \times \frac{5}{13} = \frac{45}{728}$

(II) Beginning with Black:

The probability of drawing a black ball $= \dfrac{6}{16}$

The probability of drawing a white ball then $= \dfrac{10}{15}$

The probability of drawing a black ball then $= \dfrac{5}{14}$

The probability of drawing a white ball then $= \dfrac{9}{13}$

The probability of the compound even $= \dfrac{6}{16} \times \dfrac{10}{15} \times \dfrac{5}{14} \times \dfrac{9}{13} = \dfrac{45}{728}$

Therefore, the required probability $= \dfrac{45}{728} \times \dfrac{45}{728} = 0.0038$

Example 43: *In a certain college, the student engage in different sports in the following proportions:*

Foot ball : 60 *percent of all students.*

Cricket : 50 *percent of all students.*

Both Football and Cricket: 30 *percent of all students. If student is selected at random. What is the probability that he/she will:*

(i) Play football or cricket; (ii) Play neither sport. **[B.C.A. (Meerut) 2006]**

Solution: Let number of total students $= 100$

Played football $= 60$

Played cricket $= 50$

Both played $= 30$

Played only football $= 30$

Played only cricket $= 20$

Not played $= 100 - (30 + 30 + 20) = 100 - 80 = 20$

Probability to play football $= \dfrac{30}{100} = \dfrac{3}{10}$

or Probability to play cricket $= \dfrac{20}{100} = \dfrac{1}{5}$

Probability to either sport $= \dfrac{20}{100} = \dfrac{1}{5}$

Example 44: *A bag contains 7 white, 5 black and 4 red balls. Four balls are drawn without replacement. Find the probability that at least three balls are black.* **[B.C.A. (Rohilkhand) 2008]**

Solution: Total number of black balls $= 5$

Total number of non-black balls $= 7 + 4 = 11$

$$\text{Required probability} = P(5 \text{ black balls}) + P(4 \text{ black balls})$$

$$+ P\ (3 \text{ black balls}) + P(2 \text{ black balls})$$

$$= \left(\frac{5}{16}\times\frac{4}{15}\times\frac{3}{14}\times\frac{11}{3}\right) + \left(\frac{5}{16}\times\frac{4}{15}\times\frac{11}{14}\times\frac{3}{13}\right)$$

$$+ \left(\frac{5}{16}\times\frac{11}{15}\times\frac{4}{14}\times\frac{3}{13}\right) + \left(\frac{11}{16}\times\frac{5}{15}\times\frac{4}{14}\times\frac{3}{13}\right)$$

$$= \frac{660}{43{,}680} + \frac{660}{43{,}680} + \frac{660}{43{,}680} + \frac{660}{43{,}680} = \frac{2{,}640}{43{,}680}$$

$$P(4 \text{ black balls}) = \frac{5}{16}\times\frac{4}{15}\times\frac{3}{14}\times\frac{2}{13} = \frac{120}{43{,}680}$$

$$\text{Hence, the required probability } = \frac{2{,}640}{43{,}680} + \frac{120}{43{,}680}$$

$$= \frac{2{,}760}{43{,}680} = \frac{23}{364}$$

Example 45:

(i) *In a simultaneaus toss of four coins what is the probability of getting exactly three heads ?*

(ii) *Given two mutually exclusive events A and B such that* $P(A) = 0.45$ *and* $P(B) = 0.35$, $P(A \text{ or } B)$

(iii) *In a single throw of two dice, find the probability of getting a total* 9 *or* 11.

(iv) *Two dice are thrown. Find the probability of getting an odd number on one and multiple of three on the other.*

(v) *A bag contains 5 green and 7 red balls. Two balls are drawn. What is the probability that one is green and other is red ?*

(vi) *A bag contains 5 red, 5 white and 4 black balls. Two balls are drawn. Find the probability that one of them is red.* **[B.C.A. (Agra) 2009]**

(vii) *The probability of hitting a target by three marks man are* $\frac{1}{2}, \frac{1}{3}$ *and* $\frac{1}{4}$ *respectively. Find the probability that one and only of them will hit the target when they fire simultaneously.* **[B.C.A. (Lucknow) 2008, 2004]**

(viii) *A box x contains* 3 *white balls and* 2 *black balls, and another bag y contains* 2 *white balls and* 4 *black balls. A bag and a ball out of it are picked at random. What is the probability that ball is white.* **[B.C.A. (Meerut) 2005]**

(ix) *A and B throw a coin alternatively till one of them gets a head and wins the game. Find the probability of winning.* **[B.C.A. (Kurukshetra) 2006]**

Solution:

(i) We have

S = {*HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTTH, TTHH, THHT, HTHT, THTH, HTTH, THTT, TTHT, TTTH, HITTT*}

$$P(3 \text{ heads}) = \frac{4}{16}$$

(ii) Since A and B are mutually exclusive events.

$$\therefore \quad P(A \text{ or } B) = P(A) + P(B) = 0.45 + 0.25 = 0.70$$

(iii) Let A = {getting a total of 9}

B = {getting a total of 11}

Favourable ways for getting total of 9 are: (6, 3), (3, 6), (5, 4), (4, 5)

Favourable ways for getting total of 11 are (6, 5), (5, 6)

Here, A and B are mutually exclusive events.

$$\therefore \quad P(A \text{ or } B) = P(A) = P(B) = \frac{4}{36} + \frac{2}{36} = \frac{6}{36} = \frac{1}{6}$$

(iv) Favourable ways are:

$(1,3),(1,6),(3,1),(3,3),(3,5),(3,6),(5,3),$

$(5,6),(6,1),(6,3),(6,5) =$ 11 ways

Total no. of ways $= 6 \times 6 = 36$ ways

$$\text{Required probability} = \frac{11}{36}$$

(v) $$\text{Required probability} = \frac{5\,C_1 \times 7C_1}{12\,C_2} = \frac{5 \times 7}{\frac{12 \times 11}{2 \times 1}} = \frac{35}{66}$$

(vi) Total balls $= 6 + 5 + 4 = 15$

Two balls can be drawn $= 15\,C_2$ ways

Favourable ways of getting 2 non red balls $= 9C_2$

$$\therefore \quad \text{Required probability} = \frac{9C_2}{15C_2} = \frac{\frac{9 \times 8}{2 \times 1}}{\frac{15 \times 14}{2 \times 1}} = \frac{36}{105} = \frac{12}{35}$$

(vii) Here, $$P(A) = \frac{1}{2}, P(B) = \frac{1}{3}, P(C) = \frac{1}{4}$$

and $$P(\overline{A}) = \frac{1}{2}, P(\overline{B}) = \frac{2}{3}, P(\overline{C}) = \frac{3}{4}$$

$$\therefore \quad \text{Required probability} = P(A)\,P(\overline{B})\,P(\overline{C}) + P(\overline{A})\,P(B)\,P(\overline{C}) + P(\overline{A})\,P(\overline{B})\,P(C)$$

$$= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} + \frac{1}{2} \times \frac{1}{3} \times \frac{3}{4} + \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4}$$

$$= \frac{1}{4} + \frac{1}{8} + \frac{1}{12} = \frac{6 + 3 + 2}{24} = \frac{11}{24}$$

(viii) $$\text{Probability for choosing bag } X = \frac{1}{2}$$

Probability of getting white ball out of it $= \dfrac{3}{5}$

$\therefore$ Probability to get a white ball out of X bag

$$= \frac{1}{2} \times \frac{3}{5} = \frac{3}{10}$$

Similarly probability to get a white ball out of Y bag

$$= \frac{1}{2} \times \frac{2}{6} = \frac{1}{6}$$

$\therefore$ Required probability $= \dfrac{3}{10} + \dfrac{1}{6} = \dfrac{9+5}{30} = \dfrac{14}{30} = \dfrac{7}{15}$

(ix) $P(H) = \dfrac{1}{2}, \; P(T) = \dfrac{1}{2}$

A will win if he throws head in Ist, 3rd, 5th, ...throws.

$\therefore$ A's probability of winning

$$= \frac{1}{2} + \left(\frac{1}{2}\right)^2 . \frac{1}{2} + \left(\frac{1}{2}\right)^4 . \frac{1}{2} + \ldots$$

$$= \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3} \quad \left[\therefore \; S_\infty = \frac{9}{1-r}\right]$$

and B's probability of winning $= 1 - \dfrac{2}{3} = 1/3$.

Example 46:

(i) *If $E_1, E_2, \ldots, E_n$ are a set of mutually exclusive events and exhaustive set of events, prove that the non-occurrence of E_1 or E_2 or E_3 means that at least one of the remaining events $E_1, E_2, \ldots, E_n$ must have occurred.*

(ii) *What is the probability that number selected from the numbers 1,2, ...,25 is a prime number. You may assume that each of the 25 numbers is equally likely to be selected.*

[B.C.A. (Kanpur) 2006, 2004]

(iii) *One card is drawn from a pack of 52 cards each of 52 cards being equally likely to be drawn. Find the probability of*

(a) *The card drawn is red*

(b) *The card drawn is a king*

(c) *The card drawn is red and a king*

(d) *The card drawn is either red or a king* **[B.C.A. (Kanpur) 2007]**

(iv) *One card is drawn from a pack of 52 cards so that each card is equally likely to be selected. In which of the following cases are the events E and F are independent:*

(a) *E: The card drawn is a spade*

F: The card drawn is an ace

(b) *E: The card drawn is black*

F: The card drawn is a king

(c) *E: The card drawn is a king or a queen*

F: The card drawn is a queen or jack. **[B.C.A. (Kurukshetra) 2010]**

Solution:

(i) Since $E_1, E_2 \dots, E_n$ are a set of mutuality exclusive events and exhausline set of events.

$$\therefore \qquad S = E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n$$

Where S is the sample space of experiment.

$$S = (E_1 \cup E_2 \cup E_3) \cup (E_4 \cup E_5 \cup \dots \cup E_n)$$

Let the outcome of the experiment be denoted by x since the events E_1, E_2, E_3 are given to have not occurred.

$$\therefore \qquad x \notin E_1 \cup E_2 \cup E_3.$$

$$\text{Also} \qquad x \in E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n$$

$$\therefore \qquad x \in E_4 \cup E_5 \cup \dots \cup E_n \qquad [\because\ x \in E_1 \cup E_2 \cup E_3]$$

$\Rightarrow x$ belongs to at least one out of $E_4, E_5 \dots E_n$

Hence the non-occurrence of E_1 or E_2 or E_3 means that at least one of the remaining events $E_4, E_5 \dots E_n$ must have occurred.

(ii) Here, $S = \{1,2,3,\dots,25\}$

Let A be the event of getting a prime number.

$$\therefore \qquad A = \{2,3,5,7,11,13,17,19,21\}$$

$$\therefore \qquad \text{Required probability } i.e.\ P(A) = \frac{9}{25}$$

(iii) (a) There are 26 red cards in a pack of 52 cards.

$$\therefore \qquad P(\text{a red card}) = \frac{{}^{26}C_1}{{}^{52}C_1} = \frac{26}{52} = \frac{1}{2}.$$

(b) There are 4 kings in a pack of 52 cards

$$\therefore \qquad P(\text{a king}) = \frac{{}^{4}C_1}{{}^{52}C_1} = \frac{4}{52} = \frac{1}{13}.$$

(c) There are two kings which are also red cards

$$\text{Required Prob.} = \frac{{}^{2}C_1}{{}^{52}C_1} = \frac{2}{52} = \frac{1}{26}$$

(d) Let $A = \{\text{getting a red card}\}$

$B = \{\text{getting a king}\}$

Then, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$$= \frac{1}{2} + \frac{1}{13} - \frac{1}{26} = \frac{13+2-1}{26} = \frac{14}{26} = \frac{7}{13}.$$

(iv) (a) There are 13 cards of spade in a pack

$$P(E) = \frac{13}{52} = \frac{1}{4}$$

There are 4 aces in the pack

$$P(F) = \frac{4}{52} = \frac{1}{13}$$

$$E \cap F = \{\text{an ace of spade}\}$$

$$P(E \cap F) = \frac{1}{52} \qquad (\therefore\ n(E \cap F) = 1)$$

Also, $P(E) \times P(F) = \frac{1}{4} \times \frac{1}{13} = \frac{1}{52}$

Since , $P(E \cap F) = P(E) \times P(F)$

Hence, the events E and F are independent events.

(b) There are 26 black cards in the pack

$\therefore$ $P(E) = \frac{26}{52} = \frac{1}{2}$

There are 4 kings in the pack

$$P(F) = \frac{4}{52} = \frac{1}{13}$$

$$E \cap F = \{\text{2 black king}\}$$

$$P(E \cap F) = \frac{2}{52} = \frac{1}{26}$$

Also, $P(E) \times P(F) = \frac{1}{2} \times \frac{1}{13} \times \frac{1}{26}$

Since, $P(E \cap F) = P(E) \times P(F)$

Hence, the events E and F are independent events.

(c) There are 4 kings and 4 queens in the pack

$\therefore$ $n(E) = 8$

$$\therefore \qquad P(E) = \frac{8}{52} = \frac{2}{13}$$

There are 4 queens and 4 Jacks in the pack

$$\therefore \qquad n(F) = 8$$

$$\therefore \qquad P(F) = \frac{8}{52} = \frac{2}{13}$$

$$E \cap F = \{4 \text{ queens}\}$$

$$\therefore \qquad P(E \cap F) = \frac{4}{52} = \frac{1}{13}$$

Also, $P(E) \times P(F) = \frac{2}{13} \times \frac{2}{13} = \frac{4}{169}$

Here, $P(E \cap F) \neq P(E) \times P(F)$

Hence the events E and F are not independent.

Example 47: *Three persons A, B, C throw a dice in succession till one gets a 'six' & wins the game. Find their respective probabilities of winning.* **[B.C.A. (A.M.U.) 2011, 2009]**

Solution: The chance of getting a 'six' when a dice is thrown $= \frac{1}{6}$

$\therefore$ The chance of not getting a 'six', when a dice is not thrown $= 1 - \frac{1}{6} = \frac{5}{6}$

The prob. of A' 's' winning

$$= \frac{1}{6} + \left(\frac{5}{6}\right)^3 \times \frac{1}{6} + \left(\frac{5}{6}\right)^6 \times \frac{1}{6} + \ldots\ldots$$

$$= \frac{\frac{1}{6}}{1 - \left(\frac{5}{6}\right)^3} = \frac{\frac{1}{6}}{1 - \frac{125}{216}} = \frac{\frac{1}{6}}{\frac{91}{216}} = \frac{1}{6} \times \frac{216}{91} = \frac{36}{91}$$

The prob. of B's winning

$$= \frac{5}{6} \times \frac{1}{6} \times \left(\frac{5}{6}\right)^4 \times \frac{1}{6} + \left(\frac{5}{6}\right)^7 \times \frac{1}{6} + \ldots.$$

$$= \frac{\frac{5}{6} \times \frac{1}{6}}{1 - \left(\frac{5}{6}\right)^3} = \frac{\frac{5}{36}}{1 - \frac{125}{216}} = \frac{\frac{5}{36}}{\frac{91}{216}}$$

$$= \frac{5}{36} \times \frac{216}{91} = \frac{5 \times 6}{91} = \frac{31}{90}$$

The prob. of C's winning

$$= \left(\frac{5}{6}\right)^2 \times \frac{1}{6} + \left(\frac{5}{6}\right)^5 \times \frac{1}{6} + \left(\frac{5}{5}\right)^8 \times \frac{1}{6} + \ldots.$$

$$= \frac{\left(\frac{5}{6}\right)^2 \times \frac{1}{6}}{1 - \left(\frac{5}{6}\right)^3} = \frac{\frac{25}{36 \times 6}}{1 - \frac{125}{216}} = \frac{\frac{25}{216}}{\frac{91}{216}}$$

$$= \frac{25}{91}.$$

Example 48: *A bag contains 3 red and 5 black balls and a second bag contains 6 black balls and 4 red balls. A ball is drawn from each bag. Find probability that one is red and the other is black.*

[B.C.A. (Kashi Vidhyapeeth) 2011, 2007]

Solution: Total no. of balls in the Ist bag $= 3 + 5 = 8$

$$P \text{ (getting red ball)} = \frac{3}{8}$$

and $$P \text{ (getting black ball)} = \frac{5}{8}$$

Total no. of balls in 2nd bag $= 6 + 4 = 10$

$$P \text{ (getting black ball)} = \frac{6}{10} = \frac{3}{5}$$

$$P\text{(getting red ball)} = \frac{4}{10} = \frac{2}{5}$$

Required probability = P (red ball from Ist bag) $\times P$ (black ball from 2nd bag)

+ P (red from 2nd bag) $\times P$ (black from Ist bag)

$$= \frac{3}{8} \times \frac{3}{5} + \frac{2}{5} \times \frac{5}{8} = \frac{19}{40}$$

Problem Set

1. If two balls are drawn from a bag containing 2 white, 4 red and 5 black balls. What is the chance that (*i*) both the balls are red (*ii*) one is red and the other black ?

$$\left[\textbf{Hint: } \text{(i) } \frac{^4C_2}{^{11}C_2} = \frac{6}{55}; \quad \text{(ii) } \frac{4\times 5}{^{11}C_2} = \frac{4}{11}\right]$$

2. A box contains 6 red, 4 white and 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls drawn. There is at least one ball of each colour.

[**Hint:** The required probability is

$$\frac{^6C_1 \times {^4C_1} \times {^5C_2}}{^{15}C_4} + \frac{^6C_2 \times {^4C_1} \times {^5C_1}}{^{15}C_4} + \frac{^6C_1 \times {^4C_2} \times {^5C_1}}{^{15}C_4} = \frac{48}{91}\Bigg]$$

3. What is the chance that a non-leap year selected at random will contain 53 Sunday?

[**Hint:** In a non-leap year there are 52 weeks and 1 day (which can be any day of the week). There will be 53 Sundays if the remaining one day in Sunday whose probability is $\frac{1}{7}$].

4. From a pack of 52 cards, two are drawn at random. Find the chance that one is a king and the other a queen.

$$\left[\textbf{Hint: } \frac{^4C_1 \times {^4C_1}}{^{52}C_2} = \frac{8}{663}\right].$$

5. The chance of winning of two race horses are 1/3 and 1/6 respectively. What is the probability that at least one will win when the horses are running (*i*) in different races and (*ii*) in the same race ?

$$\left[\textbf{Hint: } \text{(i) Reqd. Prob.} = 1 - \left[\left(1-\frac{1}{3}\right)\left(1-\frac{1}{6}\right)\right] = \frac{4}{9} \text{ (ii) Reqd. Prob.} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}\right]$$

6. Three ships *A*, *B* and *C* sail from England to India. Odds in favor of their arriving safely are 2 : 5, 3 : 7 and 6 : 11 respectively. Find the chance that they all arrive safely.

$$\left[\textbf{Hint: } \text{(i) Reqd. Prob.} = \frac{2}{7} \times \frac{3}{10} \times \frac{6}{17} = \frac{18}{595}\right]$$

7. From a group of 8 children, 5 boys and 3 girls. Three children are selected at random. Calculate the probabilities that the selected group contains (i) no girl, (ii) only one girl, (iii) one particular girl, (iv) at least one girl and v more girls than boys.

$$\left[\textbf{Hint: } \text{(i) } \frac{{}^5C_3}{{}^8C_3} \qquad \text{(ii) } \frac{{}^3C_1 \times {}^5C_2}{{}^8C_3} \qquad \text{(iii) } \frac{{}^5C_3}{{}^8C_3}\right.$$

$$\text{(iv) } \left(\frac{{}^3C_1 \times {}^5C_2}{{}^8C_3}\right) + \left(\frac{{}^3C_2 \times {}^5C_1}{{}^8C_3}\right) + \left(\frac{{}^3C_3}{{}^8C_3}\right)$$

$$\left.\text{(v) } \left(\frac{{}^3C_2 \times {}^5C_1}{{}^8C_3}\right) + \left(\frac{{}^3C_3}{{}^8C_3}\right)\right]$$

8. An urn contains 3 red and 4 black balls. Two balls are taken out at random. Find the probabilities that the ball are of (i) different colours (ii) black colour, (iii) red colour.

$$\left[\textbf{Hint: } \text{(i) } \frac{{}^3C_1 \times {}^4C_1}{{}^7C_2} = \frac{4}{7} \quad \text{(ii) } \frac{{}^4C_2}{{}^7C_2} = \frac{2}{7} \quad \text{(iii) } \frac{{}^3C_2}{{}^7C_2} = \frac{1}{7}\right]$$

9. Find the probability of throwing 15 with three unbiased dice.

[**Hint:** We can throw 15 in the following ways:

(6, 6, 3), (6, 3, 6), (3, 6, 6), (6, 5, 4), (6, 4, 5),

(5, 4, 6), (5, 6, 4), (4, 5, 6), (4, 6, 5), (5, 5, 5)

$$m = 10, n = 6 \times 6 \times 6 = 216]$$

10. A person throw 3 coins. Find the probability that at least one lead turns up.

11. Find the chance of throwing more than 14 with three symmetrical dice.

[**Hint:** Reqd. Prob. $= P\,[15] + P\,[16] + P\,[17] + P\,[18]$]

12. A bag contains 4 red and 3 blue balls. Two drawings of 2 balls are made. Find the chance that the first drawing gives 2 red balls and the second drawing two blue balls (i) if the balls are returned to the bag after the first draw, (ii) If the balls are not returned

$$\left[\textbf{Hint: } \text{(i) } \left(\frac{{}^4C_2}{{}^7C_2}\right)\left(\frac{{}^3C_2}{{}^7C_2}\right) \quad \text{(ii) } \left(\frac{{}^4C_2}{{}^7C_2}\right)\left(\frac{{}^3C_2}{{}^5C_2}\right)\right]$$

Answers

7.	(i) $\frac{5}{28}$, (ii) $\frac{15}{28}$, (iii) $\frac{5}{28}$, (iv) $\frac{33}{28}$, (v) $\frac{2}{7}$			**9.**	$\frac{10}{216}$
10.	$1-\left(\frac{1}{2}\right)^3 = \frac{7}{8}$	**11.**	$\frac{5}{54}$	**12.**	$\left(\frac{2}{49}, \frac{3}{35}\right)$

7.2 Set Theoretic Concepts

Let x be any universal set (*e.g.* $X = R$. set of all real numbers). Let A and B be any two subsets of X. Then

1. $A \cup B = \{x : x \in A \text{ or } x \in B \text{ } or \text{ } x \in \text{both } A \text{ and } B\}$

   is called **union** of A and B.

2. $A \cap B$ or $AB = \{x : x \in A \text{ and } x \in B\}$

   is called the **intersection** of A and B.

3. $A - B = \{x : x \in A \text{ but } x \notin B\}$

   is called the **difference** of A and B.

4. $\overline{A}$ or $X - A = \{x : x \in X \text{ but } x \notin A\}$

   is called the **complement** of A w.r.t. x

5. The sets A and B are said to be mutually exclusive or disjoint if $A \cap B = \phi$ or $AB = \phi$, where ϕ denotes the empty set. Further, a sequence $\langle A_i \rangle$ of subsets of X is said to be mutually exclusive if $A_i \, A_j = \phi$ for each $i \neq j$.

(A) Sample Space **[B.C.A. (Meerut) 2010]**

Suppose we toss a coin. The possible outcomes of this experiment are head (H) and tail (T). The collection of those outcomes is written as

$$A = \{H, T\}$$

Suppose we toss two coins at a time. The collection of all possible outcomes of this experiment is

$$B = \{HH, HT, TH, TT\}$$

If we toss three coins at a time, the collection of all possible outcomes of this experiment is

$$C = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

The collection A, B, C are called **sample spaces** of the experiments of tossing a coin, two coins and three coins : respectively. We call each of the above experiments as random experiment in as much as its outcomes is uncertain and a given outcome is just one of many possible outcomes.

The Sample Space is the collection of all possible outcomes of a random experiment.

(B) Probability Space: If Ω is a sample space. A a σ algebra of events (each a subset of Ω) and $P\,[.]$ a probability function with domain A. Then the triplet $(\Omega, A, P\,[.])$ is called a **probability space.**

Remarks:

(i) $P[\phi] = 0$

(ii) $P[A_1 \cup A_2 \cup \ldots \cup A_n] = \sum_{i=1}^{n} P[A_i]$

where $A_1, A_2, \ldots, A_n$ are mutually exclusive events in A

(iii) For any event A in A.P $[\overline{A}] = 1 - P[A]$

(iv) For any two events A and B in A.

$$P[A \cup B] = P[A] + P[B] - P[AB]$$

(*v*) If A and B are in A and $A \subset B$. Then $P[A] \le P[B]$.

Theorem: If A and B are mutually disjoint events, then

$$P[A \cup B] = P[A] + P[B]$$

Proof: Since A and B are mutually disjoint events, therefore

$$AB = \phi \Rightarrow P[AB] = P[\phi] = 0$$

Result follows by part (*iv*) now.

Theorem: (Bon-Ferroni's inequality)

If $A_1, A_2, \ldots, A_n \in A$. Then

$$P[A_1 A_2 \ldots A_n] \ge P[A_1] + P[A_2] + \ldots + P[A_n] - (n-1)$$

or

$$P\left[\sum_{i=1}^{n} A_i\right] \ge \sum_{i=1}^{n} P[A_i] - (n-1)$$

Example 49: *If* $P[A] = \frac{1}{3}$ *and* $P[\overline{B}] = \frac{1}{4}$ *can A and B be disjoint ? Explain.*

Solution: Suppose A and B are disjoint events. Then

$$A \cap B = \phi \Rightarrow A \subseteq \overline{B} \Rightarrow P[A] \le P[\overline{B}] \Rightarrow \frac{1}{3} \le \frac{1}{4}$$

which is impossible. Hence A and B cannot be disjoint.

Example 50: *A and B are disjoint events.* $P[A] = .5, P[A \cup B] = .6$. *What is* $P[B]$?

Solution: We know $P[A \cup B] = P[A] + P[B] - P[AB]$

Since A and B are disjoint $AB = \phi \Rightarrow [AB] = P[\phi] = 0$

Hence, $$P[B] = P[A \cup B] - P[A] = .6 - .5 = 0.1$$

Example 51: *If $B \subset A$, then*

(i) $P[A \cap \overline{B}] = P[A] - P[B]$

(ii) $P[B] \leq P[A]$

Solution:

(i) Since $B \subset A$. So B and $A \cap \overline{B}$ are mutually disjoint events and

$$A = B \cup (A \cap \overline{B}).$$

$$\Rightarrow \qquad P[A] = P[B \cup (A \cap \overline{B})] = P[B] + P[A \cap \overline{B}] \qquad \text{... (1)}$$

Hence, $P[A \cap \overline{B}] = P[A] - P[B]$

(ii) We know $P[A \cap \overline{B}] \geq 0 \Rightarrow P[A] - P[B] \geq$ by (1)

Hence, $P[B] \leq P[A]$

Corollary. Since $A \cap B \subset A$ and $A \cap B \subset B$, therefore,

$$P[A \cap B] \leq P[A] \text{ and } P[A \cap B] \leq P[B]$$

Example 52: *If A, B and C are any three events, then*

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B]$$
$$- P[B \cap C] - P[C \cap A] + P[A \cap B \cap C]$$

Solution: We have $P[A \cup B \cup C] = P[(A \cup B) \cup C]$

$$= P[A \cup B] + P[C] - P[(A \cup B) \cap C], \qquad \text{...(1)}$$

Now, $P[(A \cup B) \cap C] = P[(A \cap C) \cup (B \cap C)]$, by distributive law

$$= P[A \cap C] + P[B \cap C] - P[(A \cap C) \cap (B \cap C)],$$

$$= P[A \cap C] + P[B \cap C] - P[A \cap B \cap C] \qquad \text{...(2)}$$

Also $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ ...(3)

Using (2) and (3) in (1) we get the required result.

Example 53: *Three newspapers A, B, C are published in a city and a survey of readers indicates the following:*

20% *read A,* 16% *read B,* 14% *read C*

8% *read both A and B,* 5% *read both A and C*

4% *read both B and C,* 2% *read all the three.*

For a person chosen at random, find the probability that he reads none of the paper.

Solution: We have $P[A] = 20\% = \dfrac{20}{100}$, $P[B] = \dfrac{16}{100}$, $P[C] = \dfrac{14}{100}$,

$$P[A \cap B] = \frac{8}{100}, P[A \cap C] = \frac{5}{100}$$

$$P[B \cap C] = \frac{4}{100}, P[A \cap B \cap C] = \frac{2}{100}$$

The probability that the person reads A or B or C is given by

$$\begin{aligned} P[A \cup B \cup C] &= P[A] + P[B] + P[C] - P[A \cap B] - P[B \cap C] \\ &\quad - P[C \cap A] + P[A \cap B \cap C] \\ &= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{5}{100} - \frac{4}{100} + \frac{2}{100} \\ &= \frac{35}{100} = \frac{7}{20} \end{aligned}$$

Hence, the probability that he reads none of the papers

$$= 1 - P[A \cup B \cup C] = 1 - \frac{7}{20} = \frac{13}{20}$$

Example 54: *If $P[A] = 0.9, P[B] = 0.8$, show that $P[AB] \geq 0.7$.*

Solution: By Bon-ferroni inequality, we have

$$\begin{aligned} P[AB] &\geq P[A] + P[B] - 1 \\ &\geq 0.9 + 0.8 - 1 = 0.7 \end{aligned}$$

Hence, $P[AB] \geq 0.7$

Example 55: *Event A and B are such that $P[A] = \frac{3}{4}$ and $P[B] = \frac{5}{8}$, show that*

(*i*) $P[A \cup B] \geq \frac{3}{4}$

(*ii*) $\frac{3}{8} \leq P[AB] \leq \frac{5}{8}$

(*iii*) $\frac{1}{8} \leq P[A\overline{B}] \leq \frac{3}{8}$

Solution:

(i) Since $A \subseteq A \cup B, P[A] \leq P[A \cup B]$

Thus, $\frac{3}{4} \leq P[A \cup B] \Rightarrow P[A \cup B] \geq \frac{3}{4}$

(ii) Since $AB = A \cap B \subseteq B \Rightarrow P[AB] \leq P[B] = \frac{5}{8}$

Thus, $P[AB] \leq \frac{5}{8}$, by Bon-ferroni's inequality

$$P[AB] \geq P[A] + P[B] - 1 = \frac{3}{4} + \frac{5}{8} - 1 = \frac{3}{8}$$

$\therefore$ $\frac{3}{8} \leq P[AB] \leq \frac{5}{8}$

(iii) Again $P\,[A\overline{B}] \geq P\,[A] + P\,[\overline{B}] - 1 = \frac{3}{4} + \left(1 - \frac{5}{8}\right) - 1 = \frac{1}{8}$

Since $A\overline{B} = A \cap \overline{B} \subseteq \overline{B}\,.\, P\,[A\overline{B}] \leq P\,[\overline{B}] = 1 - \frac{5}{8} = \frac{3}{8}$

$\therefore$ $\frac{1}{8} \leq P\,[A\overline{B}] \leq \frac{3}{8}$

Example 56: *Events A and B are such that*

$$P\,[A \cup B] = \frac{3}{4}, P\,[AB] = \frac{1}{4} \text{ and } P\,[\overline{A}] = \frac{2}{3}$$

Find $P\,[B]$ and $P\,[A\overline{B}]$

Solution:

We know $P\,[A \cup B] = P\,[A] + P\,[B] - P\,[AB]$

$\Rightarrow$ $\frac{3}{4} = \left(1 - \frac{2}{3}\right) + P\,[B] - \frac{1}{4} \Rightarrow P\,[B] = \frac{2}{3}$

Now, $P\,[A\overline{B}] = P\,[A \cap \overline{B}] = P\,[A] - P\,[AB]$,

$\therefore$ $P\,[A\overline{B}] = \left[1 - \frac{2}{3}\right] - \frac{1}{4} = \frac{1}{12}$

Example 57: *In a very hotly fought battle, the probability of a combatant losing an eye was 70%, an ear 75%, an arm 80% and a leg 85%. Show that the probability of a combatant losing all the four was not less than 10%.*

Solution: Let A_1, A_2, A_3 *and* A_4 denote the events of the combatant losing an eye, an arm, an ear, and a leg respectively. Then

$$P\,[A_1] = 0.70, P\,[A_2] = 0.75, P\,[A_3] = 0.80, P\,[A_4] = 0.85$$

By Bon-ferroni's inequality

$$P\,[A_1\ A_2\ A_3\ A_4] \geq P\,[A_1] + P\,[A_2] + P\,[A_3] + P\,[A_4] - (4 - 1)$$

$$\geq 0.70 + 0.75 + 0.80 + 0.85 - 3 = 0.10$$

$\therefore$ $P\,[A_1\ A_2\ A_3\ A_4] \geq 0.10$

$\Rightarrow$ $P\,[A_1\ A_2\ A_3\ A_4] \nless 0.10\,(10\%)$

Example 58: *A purse contains 2 silver and 4 copper coins. A second purse contains 4 silver and 3 copper coins. If a coin is pulled out at random from one of the two purses, what is the probability that it is a silver coin ?*

[B.C.A. (B.H.U.) 2010]

Solution: Probability of choosing first purse $= \frac{1}{2}$

Probability to getting a silver coin out of it $= \frac{2}{6} = \frac{1}{3}$

∴ Probability to get a silver coin and first purse

$$=\frac{1}{2}\times\frac{1}{3}=\frac{1}{6}$$

Similarly probability to get a silver coin and IInd purse

$$=\frac{1}{2}\times\frac{4}{7}=\frac{2}{7}$$

∴ Required probability $=\frac{1}{6}+\frac{2}{7}=\frac{2+12}{42}=\frac{19}{42}$.

Example 59: *One bag contains five white and four black balls. Another bag contains seven white and nine black balls. A ball is transferred from the first bag to second bag and then a ball is drawn from the second. Find the probability that the ball drawn is white.*

[B.C.A. (Kashi Vidhyapeeth) 2009]

Solution: The ball transferred from the first bag to second bag can either be white or black.

Case I: When white ball is transferred from first bag to IInd bag. Probability of getting white ball from Ist bag.

$$=\frac{5}{(5+4)}=\frac{5}{9}$$

Now second bag contains 8 white and 9 black balls. The probability of drawing a white ball from the IInd bag

$$=\frac{8}{(8+9)}=\frac{8}{17}$$

∴ The probability of both these events taking place together is

$$=\frac{5}{9}\times\frac{8}{17}=\frac{40}{153}.$$

Case II: When a black ball is transferred from Ist bag to IInd bag.

Probability of getting black ball from Ist bag

$$=\frac{4}{(5+4)}=\frac{4}{9}.$$

Now, second bag contains 7 white and 10 black balls. The probability of drawing a white ball from second bag

$$=\frac{7}{(7+10)}=\frac{7}{17}.$$

∴ The probability of both these events taking place together

$$=\frac{4}{9}\times\frac{7}{17}=\frac{28}{153}$$

Since the above two case are mutual exclusive, therefore only of them can happen. Hence the required probability

$$=\frac{40}{153}+\frac{28}{153}=\frac{68}{153}=\frac{4}{9}.$$

Example 60: *An almirah stores 5 black and 4 white balls. A boy pulls out 2 balls at random. What is the probability of the same colour ?* **[B.C.A. (Meerut) 2002]**

Solution: Total no. of balls $= 5 + 4 = 9$

$\therefore$ Total no. of ways choosing 2 balls out of 9 balls

$$N(S) = {}^9C_2 = \frac{\underline{|9}}{\underline{|9-2}\ \underline{|2}} = 36$$

No. of favourable events $N(E) = {}^5C_2 + {}^4C_2 = 16$

$\therefore$ P(Getting 2 balls of same colour) $= \dfrac{N(E)}{N(S)} = \dfrac{16}{36} = \dfrac{4}{9}$

Example 61: *A man has 3 children with atleast one boy. What is the probability that he has 2 boys and one girl.* **[B.Ed. Entrance Exam.-2002]**

Solution: Total no. of events $N(S)$

$$= {}^3C_1 \times {}^3C_2 + {}^3C_2 \times {}^3C_1 = 3 \times 3 + 3 \times 3 = 18$$

Favourable no. of events $N(E) = {}^3C_2 \times {}^3C_1 = 3 \times 3 = 9$

$\therefore$ P (getting 2 boys and one girl) $= \dfrac{N(E)}{N(S)}$

$\Rightarrow$ $$P(E) = \frac{9}{18} = \frac{1}{2}$$

Example 62: *There are six different letters and 6 correspondingly addressed envelops. If the letters are randomly put in the envelopes, what is the probability that exactly 5 letters go into the correctly addressed envelopes ?*

Solution: Probability of r particular letters out of n letters in the right envelopes

$$= \frac{\underline{|n-r}}{\underline{|n}}$$

$\therefore$ Probability that exactly 5 letters go into the correctly addressed envelopes

$$= \frac{\underline{|6-5}}{\underline{|6}} = \frac{\underline{|1}}{\underline{|6}} = \frac{1}{720} = 0.0013 \approx 0$$

7.2.1 Computation of the Probability of an Event by the Axiomatic Approach

Consider the probability space S where the sample space. S is finite. Let S contain N sample points $W_1, W_2, \ldots, W_N$. Suppose that these sample points are equally likely *i.e.*

$$P[\{W_1\}] = P[\{W_2\}] = \ldots = P[\{W_N\}] \qquad \ldots(1)$$

Then the probability of any event A is given by

$$P[A] = \frac{N(A)}{N(S)} \qquad \ldots(2)$$

Here $N(A)$ and $N(S)$ denote the number of points in A and S respectively.

The probability function $P[\,.\,]$. Satisfying conditions (1) and (2) above is said to be an equally likely probability function.

Example 63: *Find the probability of getting (i) a total of 7 and (ii) a total of 9 when two dice are thrown together.*

Solution: The sample space is

$$S = \{i, j) : 1 \le i \le 6, 1 \le j \le 6\}, \;\; N(S) = 36$$

It is easy to verify that all sample points are equally likely, as the probability of each sample point is 1/36.

Let A_7 = event that the total is 7.

A_9 = event that the total is 9.

$\therefore$ $A_7 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

$A_9 = \{(3,6), (4,5), (5,4), (6,3)\}$

We have $N(A_7) = 6, N(A_9) = 4$

Hence,

$$P[A_7] = \frac{N(A_7)}{N(S)} = \frac{6}{36} = \frac{1}{6},$$

$$P[A_9] = \frac{N(A_9)}{N(S)} = \frac{4}{36} = \frac{1}{9}$$

Example 64: *One urn contains three red balls, two white balls and one blue ball, A second urn contains one red ball, two white balls and three blue balls, one ball is selected at random from each urn.*

(i) Describe the sample space for this experiment.

(ii) If the balls in the two urns are mixed in a single urn and then a sample of three is drawn. Find the probability that all the three colours are represented when sampling is done (a) with replacement, (b) without replacement.

Solution: Urn $I = \{r_1, r_2, r_3, w_1, w_2, b\}$, Urn $II = \{R_1, W_1, W_2, B_1, B_2, B_3\}$. Here r_i, R denote red colour; b_i, B_i blue colour and w_i, W_i white colour.

(i) As each urn contains 6 balls, the sample space of the experiment consists of $6 \times 6 = 36$ points;

$$S = \{r_1 R, r_2 R; w_1 W_1, w_1 W_2, w_2 W_1, w_2 W_2; bB_1, bB_2, bB_3; \ldots\}$$

(ii) If the balls in the two urns are mixed in a single urn, then the latter urn contains 12 balls out of which 4 balls are red, 4 white and 4 blue. Draw a sample of 3 balls, where the sampling is done with replacement. The probability of the first ball (of any colour) is 12/12. Now we have 8 balls of the remaining two colours and the probability of the second ball (of the second colour) is 8/12. Lastly, the probability of the third ball (of the third colour) is 4/12. Hence, the probability that a sample of 3 balls contains all the three colours is

$$\frac{12}{12} \times \frac{8}{12} \times \frac{4}{12} = \frac{2}{9}$$

When the sampling is done without replacement the probability that a sample of three balls contains all the three colour is

$$\frac{{}^4C_1 \times {}^4C_1 \times {}^4C_1}{{}^{12}C_3} = \frac{4 \times 4 \times 4 \times 3 \times 2}{12 \times 11 \times 10} = \frac{16}{55}$$

Example 65: *An urn contains five balls numbered 1 to 5 of which the first three are black and the last two are gold. A sample of size 2 is drawn with replacement. Let B_1 denote the event that the first ball drawn is black and B_2 denote the event that the second ball drawn is black.*

(i) Describe a sample space for the experiment and exhibit the events B_1, B_2 and $B_1 B_2$.

(ii) Find $P[B_1], P[B_2]$ and $P[B_1 B_2]$

(iii) Repeat parts (i) and (ii) for sampling without replacement.

Solution: Urn $= \{b_1, b_2, b_3, g_4, g_5\}$ (b = black, g = gold)

(*i*) The sample space consists of $5 \times 5 = 25$ points viz.

$$S = \{b_1 b_1, b_1 b_2, b_1 b_3, b_1 g_4, b_1 g_5, b_2 b_1, b_2 b_2, b_2 b_3, b_2 g_4, b_2 g_5; \ldots; g_5 b_1, g_5 b_2, g_5 b_3, g_5 g_4, g_5 g_5\}$$

(*ii*) The events B_1, B_2 and $B_1\ B_2$ are described below:

$$B_1 = \{b_1 b_1, b_1 b_2, b_1 b_3, b_1 g_4, b_1 g_5, b_2 b_1, b_2 b_2, b_2 b_3, b_2 g_4, b_2 g_5, b_3 b_1, b_3 b_2, b_3 b_3, b_3 g_4, b_3 g_5\}$$

$$B_2 = \{b_1 b_1, b_1 b_2, b_1 b_3, b_2 b_1, b_2 b_2, b_2 b_3, b_3 b_1, b_3 b_2, b_3 b_3, g_4 b_1, g_4 b_2, g_4 b_3, g_5 b_1, g_5 b_2, g_5 b_3\}$$

$$B_1 B_2 = B_1 \cap B_2 = \{b_1 b_1, b_1 b_2, b_1 b_3, b_2 b_1, b_2 b_2, b_2 b_3, b_3 b_1, b_3 b_2, b_3 b_3\}$$

$$\therefore \quad P[B_1] = \frac{15}{25} = \frac{3}{5}, P[B_2] = \frac{15}{25} = \frac{3}{5}, P[B_1 B_2] = \frac{9}{25}$$

(*iii*) In sampling without replacement, the sample space S consists of $5 \times 4 = 20$ points. We obtain these 20 points by excluding the points $b_1 b_1, b_2 b_2, b_3 b_3, g_4\, g_4$ and

$g_5\ g_5$ from S. The events B_1 and B_2 now (consist of 12 points each (exclude b_1b_1, b_2b_2 and b_3b_3). Similarly B_1B_2 consists of 6 points.

Hence, $$P\,[B_1] = \frac{12}{20} = \frac{3}{5} = P\,[B_2] \text{ and } P\,[B_1B_2] = \frac{6}{20} = \frac{3}{10}$$

Example 66: *Find the probability that a certain 13-card hand contains exactly 8 spades.*

Solution: Total number of cards is $M = 52$

Total number of spades $= 13$

We can regard a 13-card hand as a sample of size 13 drawn without replacement from the 52 cards. Let A_8 denote the event of drawing exactly 8 spades in a 13-card hand. We can draw 13 cards from a pack of 52 cards in $\binom{52}{13}$ ways. 8 spades out of 13 can be selected in $\binom{13}{8}$ ways and the remaining 13-8 cards out of a total of 52-13 can be selected in $\binom{52-13}{13-8} = \binom{39}{5}$ ways.

Hence, $$P\,(A_8) = \binom{13}{8}\binom{39}{5} \div \binom{52}{13}$$

7.3 Conditional Probability

Let A and B be two events in a sample space (S). The **conditional probability** of event A given event B, denoted by $P\left[\frac{A}{B}\right]$ is defined by

$$P\left[\frac{A}{B}\right] = \frac{P\,[AB]}{P\,[B]}, \text{ if } P\,[B] > 0$$

Similarly, $$P\left[\frac{B}{A}\right] = \frac{P\,[AB]}{P\,[A]}, \text{ if } P\,[A] > 0$$

is the **conditional probability** of event B given event A. From the above relations, we see that

$$P\,[AB] = P\,[A/B]\;P\,[B] = P\left[\frac{B}{A}\right] P\,[A]$$

where $P\,[A] > 0$ and $P\,[B] > 0$.

Example 67: *Given $P\,[A] = 0.5$ and $P\,[A \cup B] = 0.6$ Find $P\,[B]$ if :*

(i) *A and B are mutually exclusive.*

(ii) *A and B are independent.*

(iii) $P\left[\frac{A}{B}\right] = 0.4$

Solution: We know $P[A \cup B] = P[A] + P[B] - P[AB]$ ...(1)

(i) If A and B are mutually exclusive $AB = \phi$

$$\therefore \qquad P[A \cup B] = P[A] + P[B] - P[\phi], \text{ where } P[\phi] = 0$$

Hence , $P[B] = P[A \cup B] - P[A] = 0.6 - 0.5 = 0.1$

(ii) If A and B are independent, then

$$P[AB] = P[A]\,P[B]$$

Putting in (1), we get

$$p[A \cup B] = P[A] + P[B] - P[A]\,P[B]$$

or $$P[A \cup B] = P[A] + P[B](1 - P[A])$$

or $$0.6 = 0.5 + P[B](1 - 0.5)$$

Hence, $$P[B] = \frac{1}{5}$$

(iii) We know $$P[A/B] = \frac{P[AB]}{P[B]} \text{ or } P[AB] = 0.4\,P[B]$$

Putting in (1), we get

$$0.6 = 0.5 + P[B] - 0.4\,P[B] = 0.5 + 0.6\,P[B]$$

Hence, $$P[B] = \frac{1}{6}$$

Example 68: *If* $P[B] = P\left[\frac{A}{B}\right] = P\left[\frac{C}{AB}\right] = \frac{1}{2}$, *what is* $P[ABC]$ *?*

Solution: By multiplication rule, we have

$$P[ABC] = P[A]\,P\left[\frac{B}{A}\right] P\left[\frac{C}{AB}\right] \qquad \text{...(1)}$$

By the definition of conditional probability we have

$$P[AB] = P\left[\frac{A}{B}\right] P[B] = P\left[\frac{B}{A}\right] P[A] \qquad \text{...(2)}$$

From (1) and (2), we have

$$P[ABC] = P\left[\frac{A}{B}\right] P[B]\,P\left[\frac{C}{AB}\right] = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$$

Example 69: *A single die is tossed; then n coins are tossed, where n is the number shown on the die. What is the probability of exactly two heads ?*

Solution: Let $B_1, B_2 \ldots B_6$ be six mutually disjoint events, where $B_j : j$ appears on the die. $1 \le j \le 6$. Clearly $$P[B_j] = \frac{1}{6}, 1 \le j \le 6.$$

Let $$B = \bigcup_{j=1}^{6} B_j$$

Let A denote the event when the head appears exactly twice, we have to find $P\,[A]$. By the theorem of total probability, we have

$$P\,[A]=\sum_{j=1}^{6} P\,[B_j]\,P\left[\frac{A}{B_j}\right]=\frac{1}{6}\sum_{i=1}^{6} P\left[\frac{A}{B_j}\right]$$

Note that if face 1 appears on the die we can not get two heads and so $P\left[\frac{A}{B_1}\right]=0.$

Thus, $$P\,[A]=\frac{1}{6}\left(0+\frac{{}^2C_2}{2^2}+\frac{{}^3C_2}{2^3}+\frac{{}^4C_2}{2^4}+\frac{{}^5C_2}{2^5}+\frac{{}^6C_2}{2^6}\right)$$

$$=\frac{1}{24}\left(1+\frac{3}{2}+\frac{6}{4}+\frac{10}{8}+\frac{15}{16}\right)$$

$$=\frac{1}{24}\left(\frac{16+24+24+20+15}{16}\right)=\frac{99}{384}$$

Example 70: *Five percent of the people have high blood pressure, of the people with high blood pressure, 75 percent drink alcohol; whereas only 50 percent of the people without high blood pressure drink alcohol. What percent of the drinkers have high blood pressure ?*

Solution: Let A denote the event that people have high blood pressure and B denote the people who drink alcohol.

We have $$P\,[A]=0.05\,,P\left[\frac{B}{A}\right]=0.75,P\left[\frac{B}{\bar{A}}\right]=0.5$$

We have to find $$P\left[\frac{B}{A}\right]=\frac{P\,[A\cap B]}{P\,[B]} \qquad ...(1)$$

We know $$B=AB\cup\bar{A}B \text{ and } AB\cap\bar{A}B=\phi$$

$\therefore$ $$P\,[B]=P\,[AB]+P\,[\overline{A}B]=P\,[A\cap B]+P\,[B\cap\bar{A}] \qquad ...(2)$$

Now $$P\left[\frac{B}{A}\right]=0.75$$

$\Rightarrow$ $$\frac{P\,[B\cap A]}{P\,[A]}=0.75$$

$\Rightarrow$ $$P\,[A\cap B]=0.75\times 0.05=0.0375 \qquad ...(3)$$

Also $$P\left[\frac{B}{\bar{A}}\right]=0.75$$

$\Rightarrow$ $$\frac{P\,[B\cap\bar{A}]}{P\,[\bar{A}]}=0.5$$

$\Rightarrow$ $$P\,[B\cap\bar{A}]=\frac{1}{2}\,\{1-P\,[A]\}$$

$\Rightarrow$ $$P\,[B]-P\,[A\cap B]=\frac{1}{2}\,(1-0.05)=\frac{1}{2}\,(0.95)=0.475, \text{ by (2)}$$

$\therefore$ $$P\,[B]=0.0375+0.475=0.5125, \text{by (3)}$$

Hence, by (1), $$P\left[\frac{A}{B}\right]=\frac{0.0375}{0.5125}=\frac{375}{5125}=0.073$$

Hence, the required percentage is 7.3%.

7.4 Baye's Theorem

Let $B_1, B_2 \ldots B_n$ be a collection of mutually disjoint events in T of the given probability space (S). Such that

$$S = \bigcup_{j=1}^{n} B_j \text{ and } P\,[B_j] > 0 \text{ for } j = 1, 2, \ldots, n$$

Then for each $A \in T$ satisfying $P\,[A] > 0$, we have

$$P\left[\frac{B_K}{A}\right] = \frac{P\left[\dfrac{A}{B_K}\right] P\,[B_K]}{\displaystyle\sum_{j=1}^{n} P\left[\frac{A}{B_j}\right] P\,[B_j]}$$

[This is known as Baye's formula]

Proof: By the definition of conditional probability, we have

$$P\left[\frac{B_k}{A}\right] = \frac{P\,[B_k A]}{P\,[A]}, \qquad (\because P\,[A] > 0) \qquad \ldots(1)$$

and

$$P\left[\frac{A}{B_k}\right] = \frac{P\,[AB_k]}{P\,[B_k]} \qquad (\because P\,[B_k] > 0) \qquad \ldots(2)$$

From (1) and (2), we obtain

$$P\left[\frac{B_k}{A}\right] = \frac{P\left[\dfrac{A}{B_k}\right] P\,[B_k]}{P\,[A]} \qquad \ldots(3)$$

By the theorem of total probabilities, we have

$$P\,[A] = \sum_{j=1}^{n} P\left[\frac{A}{B_j}\right] P\,[B_j], \text{ for each } A \in T$$

Substituting in (3), Baye's formula is proved.

The following particular case of Baye's formula is quite useful.

Corollary: If $A, B \in T$ where $P\,[A] > 0$ and $0 < P\,[B] < 1$, then

$$P\left[\frac{B}{A}\right] = \frac{P\left[\dfrac{A}{B}\right] P\,[B]}{P\left[\dfrac{A}{B}\right] P\,[B] + P\left[\dfrac{A}{\overline{B}}\right] P\,[\overline{B}]}$$

Proof: By the definition of conditional probability, we have

$$P\left[\frac{B}{A}\right] = \frac{P\,[BA]}{P\,[A]}, P\left[\frac{A}{B}\right] = \frac{P\,[AB]}{P\,[B]}$$

$$\therefore \qquad P\left[\frac{B}{A}\right] = \frac{P\left[\dfrac{A}{B}\right] P\,[B]}{P\,[A]} \qquad \ldots(4)$$

By the corollary of theorem, we have

$$P\,[A] = P\left[\frac{A}{B}\right] P\,[B]\; P\left[\frac{A}{\overline{B}}\right] P\,[\overline{B}] \qquad \text{...(5)}$$

The result follows from (4) and (5).

Remark: If an event E can occur in combination with two mutually exclusive events E_1 and E_2. Then

$$P\left[\frac{E_i}{E}\right] = \frac{P\,[E_i]\,P\left[\dfrac{E}{E_i}\right]}{P\,[E_1]\,P\left[\dfrac{E}{E_1}\right] + P\,[E_2)\,P\left[\dfrac{E}{E_2}\right]}\,; \quad i = 1, 2$$

Multiplication Rule

Let $A_1, A_2, \ldots, A_n$ be arbitrary events in T of the given probability space (S) such that $P\,[A_1\,A_2 \ldots A_{n-1}] > 0$; then

$$P\,[A_1 A_2 \ldots A_n] = P\,[A_1]\,P\left[\frac{A_2}{A_1}\right] P\left[\frac{A_3}{A_1 A_2}\right] \ldots P\left(\frac{A_n}{A_1 A_2 \ldots A_{n-1}}\right]$$

Example 71: *The probability that a person can hit a target is 3/5 and the probability that another person can hit the same target is 2/5. But the first person can fire 8 shots in the time. The second person fires 10 shots. They fire together. What is the probability that the second person shoots the target ?*

Solution: Let E denote the event of shooting the target, E_1 and E_2 respectively denote the events that the first person and the second person shoot the target we are given

$$P\left[\frac{E}{E_1}\right] = \frac{3}{5} \text{ and } P\left[\frac{E}{E_2}\right] = \frac{2}{5}$$

The ratio of the shots of the first person to those of the second person in the same time is $\frac{8}{10} = \frac{4}{5}$. Thus, $P\,[E_1] = \frac{4}{5}\,P\,[E_2]$. By Baye's theorem, we get

$$P\left[\frac{E_2}{E}\right] = \frac{P\,[E_2]\,P\left[\dfrac{E}{E_2}\right]}{P\,[E_1]\,P\left[\dfrac{E}{E_1}\right] + P\,[E_2]\,P\left[\dfrac{E}{E_2}\right]}$$

$$= \frac{P\,[E_2]\,.\,\dfrac{2}{5}}{\dfrac{4}{5}\,P\,[E_2]\,.\,\dfrac{3}{5} + P\,[E_2]\,.\,\dfrac{2}{5}}$$

$$= \frac{1}{\left(\dfrac{6}{5}\right) + 1} = \frac{5}{11}$$

Example 72: *Urn A contains 2 white and 2 black balls. Urn B contains 3 white and 2 black balls. One ball is transferred from A to B and then one ball is drawn out of B. Find the chance that this ball is white. If this ball turns out to be white, find the probability that the transferred ball was white.*

Solution: Let E_1 and E_2 denote the events that the transferred ball from A to B is white and black, respectively. Let W denote the event that a white balls is drawn from B. Then

$$P[E_1] = \frac{2}{4}, P[E_2] = \frac{2}{4}, P\left[\frac{W}{E_1}\right] = \frac{4}{6}, P\left[\frac{W}{E_2}\right] = \frac{3}{6}$$

$$\therefore \quad P[W] = P[E_1]\,P\left[\frac{W}{E_1}\right] + P[E_2]\,P\left[\frac{W}{E_2}\right] \quad ...(1)$$

$$= \frac{2}{4}\cdot\frac{4}{6} + \frac{2}{4}\cdot\frac{3}{6} = \frac{7}{12}$$

By Baye's theorem, we have

$$P\left[\frac{E_1}{W}\right] = \frac{P[E_1]\,P\left[\frac{W}{E_1}\right]}{P[W]} = \frac{\left(\frac{2}{4}\right)\left(\frac{4}{6}\right)}{\frac{7}{12}} = \frac{4}{7}$$

Example 73: *An urn contains a white balls and b black balls, another contains c white and d black balls. One ball is transferred from the first urn into the second and then a ball is drawn from the latter. What is the probability that it will be a white ball ?*

Solution: We know that

$$P[E_1] = \frac{a}{a+b}, \quad P[E_2] = \frac{b}{a+b}$$

$$P\left[\frac{W}{E_1}\right] = \frac{c+1}{(c+1)+d}, \quad P\left[\frac{W}{E_2}\right] = \frac{c}{c+(d+1)}$$

$$\therefore \quad P[W] = P[E_1]\,P\left[\frac{W}{E_1}\right] + P[E_2]\,P\left[\frac{W}{E_2}\right]$$

$$= \frac{a}{a+b}\cdot\frac{c+1}{c+d+1} + \frac{b}{a+b}\cdot\frac{c}{c+d+1}$$

$$\text{Hence} \quad P[W] = \frac{ac+bc+a}{(a+b)(c+d+1)}$$

Example 74: *Three urns A_1, A_2 and A_3 contain respectively 3 red, 4 white, 1 blue, 1 red, 2 white, 3 blue, 4 red, 3 white, 2 blue balls, one urn is chosen at random and a ball is withdrawn, it is found to be red. Find the probability that it came from urn A_2.*

Solution: If A_i denotes i^{th} urn chosen and R denotes the event of withdrawing the red ball, then

$$P[A_1] = P[A_2] = P[A_3] = \frac{1}{3}$$

Now, $$P\left[\frac{R}{A_1}\right]=\frac{3}{8},P\left[\frac{R}{A_2}\right]=\frac{1}{6},P\left[\frac{R}{A_3}\right]=\frac{4}{9}$$

By Baye's theorem, we have

$$P\left[\frac{A_2}{R}\right]=\frac{P\,[A_2]\,P\left[\dfrac{R}{A_2}\right]}{\sum\limits_{i=1}^{3}P\,[A_i]\,P\left[\dfrac{R}{A_i}\right]}$$

$$=\frac{\left(\frac{1}{3}\right)\left(\frac{1}{6}\right)}{\left(\frac{1}{3}\right)\left(\frac{3}{8}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{6}\right)+\left(\frac{1}{3}\right)\left(\frac{4}{9}\right)}=\frac{1}{18}\times\frac{216}{71}=\frac{12}{71}$$

Example 75: *Urns A, B, C have the following coloured balls:*

A : 6 red, 4 white, B : 2 red, 6 white, C : 1 red, 8 white. An urn is chosen at random; a ball drawn turns out to be red. Find the chance that urn A is chosen.

[**Hint:** $$P\,[R]=P\,[A]\left[\frac{R}{A}\right]+P\,[B]\,P\left[\frac{R}{B}\right]+P\,(C)\,P\left[\frac{R}{C}\right]$$

$$=\frac{1}{3}\left[\frac{6}{10}+\frac{2}{8}+\frac{1}{9}\right]=\frac{173}{540}$$]

By Baye's theorem

$$P\left[\frac{A}{R}\right]=\frac{P\,[A]\,P\left[\dfrac{R}{A}\right]}{P\,[R]}=\left(\frac{1}{3}\cdot\frac{6}{10}\right)\div\frac{173}{540}=\frac{108}{173}$$

Example 76: *There are five urns and they are numbered 1 to 5. Each urn contains 10 balls urn i has i defective balls and $10-i$ non-defective balls $(i=1,2,3,4,5)$. An urn is selected at random and then a ball is selected at random from the selected urn. What is the probability that a defective ball will be selected? If the ball is defective, what is the probability that it come from urn 5 ?*

Solution: Let E denote the event that a defective ball is selected and E_i the event that urn i is selected; $i=1,2,3,4,5$, we have

$$P\,[E_i]=\frac{1}{5},P\left[\frac{E}{E_i}\right]=\frac{i}{10};i=1,2,3,4,5$$

By the theorem of total probabilities, the probability that a defective ball is selected is given by

$$P\,[E]=\sum_{i=1}^{5}P\,[E_i]\,P\left[\frac{E}{E_i}\right]$$

$$=\sum_{i=1}^{5}\frac{1}{5}\cdot\frac{i}{10}=\frac{1}{50}(1+2+3+4+5)=\frac{3}{10}$$

(ii) The probability that the defective ball came from urn 5 is given by

$$P\left[\frac{E_5}{E}\right] = \frac{P[E|E_5]P[E_5]}{\sum_{i=1}^{5} P[E|E_i]P[E_i]}$$

$$= \frac{\frac{5}{10}\cdot\frac{1}{5}}{\frac{3}{10}} = \frac{1}{3}$$

Note: That the urn 5 has 5 defective balls out of a total of 10.

Example 77: *An urn contains 10 balls of which 3 are black and 7 are white. The following game is played. At each trial a ball is selected at random its colour is noted and it is replaced by two additional balls of the same colour. What is the probability that black ball is selected in each of the first three trials ?*

Solution: Let B_i denote the event that a black ball is selected in the i^{th} trial. We have to find $P\,[B_1\ B_2\ B_3]$ by the multiplication rule, we have

$$P\,[B_1, B_2, B_3] = P\,[B_1]\,P\left[\frac{B_2}{B_1}\right]P\left[\frac{B_3}{B_1, B_2}\right]$$

$$= \frac{3}{10} \times \left(\frac{3+2}{10+2}\right) \times \left(\frac{3+2+2}{10+2+2}\right)$$

$$= \frac{3}{10}\cdot\frac{5}{12}\cdot\frac{7}{14} = \frac{1}{16}$$

Similar Example: *In the above example, find the probability that a white ball is selected in each of the first three trials.*

$$\left[\frac{7}{10}\cdot\frac{9}{12}\cdot\frac{11}{14}\ \textbf{Ans.}\right]$$

Example 78: *A die is thrown as long as necessary for a 6 to turn up. Given that 6 does not turn up at the first throw. What is the probability that more than four throws will be necessary ?*

Solution: Let A_i be the event that 6 does not turn up at the ith throw where $i = 1, 2, 3, 4$. Then

$$P\,[A_i] = \frac{5}{6}$$

Notice that more than four throws are needed given A_1 means that in the first four throws 6 does not turn up.

The required probability is $P\left[\frac{A_1\ A_2\ A_3\ A_4}{A_1}\right]$ using multiplication rule, we have

$$P\left[\frac{A_1\ A_2\ A_3\ A_4}{A_1}\right] = \frac{P\,[A_1]\,P\left[\frac{A_2}{A}\right]P\left[\frac{A_3}{A_1\ A_2}\right]P\left[\frac{A_4}{A_1 A_2 A_3}\right]}{P\,[A_1]}$$

$$= P\left[\frac{A_2}{A_1}\right]P\left[\frac{A_3}{A_1 A_2}\right]P\left[\frac{A_4}{A.\ A_2 A_3}\right]$$

$$= \frac{5}{6}\times\frac{5}{6}\times\frac{5}{6} = \frac{125}{216}$$

Example 79: *An insurance company insured* 2000 *scooter drivers,* 4000 *car drivers and* 6000 *truck drivers. The probability of any accident involving a scooter, car and a truck is* 0.01, 0.03 *and* 0.15 *respectively. One of the insured persons meets with an accident. What is the probability that he is a scooter driver ?*

Solution: Let E_1, E_2 and E_3 denote the events that a scooter driver and car driver, truck driver is chosen respectively. Then

$$P\,[E_1] = \frac{2000}{12000} = \frac{1}{6},\ P\,[E_2] = \frac{4000}{12000} = \frac{1}{3},\ P\,[E_3] = \frac{6000}{12000} = \frac{1}{2}$$

Let A denote the event the any of the drivers meets with an accident. The probability that a scooter driver meets with an accident is

$$P\left[\frac{A}{E_1}\right] = 0.001.\ \text{Similarly,}\ P\left[\frac{A}{E_2}\right] = 0.03\ ;\ P\left[\frac{A}{E_3}\right] = 0.15$$

The probability that the insured person is a scooter driver is given by

$$P\left[\frac{E_1}{A}\right] = \frac{P\,[E_1]\,P\left[\frac{A}{E_1}\right]}{P\,[E_1]\,P\left[\frac{A}{E_1}\right] + P\,[E_2]\,P\left[\frac{A}{E_2}\right] + P\,[E_3]\,P\left[\frac{A}{E_3}\right]}$$

$$= \frac{\frac{1}{6}(0.01)}{\frac{1}{6}(0.01) + \frac{1}{3}(0.03) + (0.15)}$$

$$= \frac{0.01}{0.01 + 2\,(0.03) + 3\,(0.15)} = \frac{0.01}{0.52} = \frac{1}{52}$$

Example 80: *Three boxes contain galvanometers, some of which are detectives. The proportions of defective in boxes A, B and C are respectively* 1/3, 2/9 *and* 1/6. *A box is selected at random and a galvanometer drawn from it. If the selected galvanometer is found to be defective. What is the probability that the box B was selected ?*

[**Hint:** Required probability is $= \dfrac{\frac{1}{3}\cdot\frac{2}{9}}{\frac{1}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{9} + \frac{1}{3}\cdot\frac{1}{6}} = \dfrac{4}{13}$]

Example 81: *In a bolt factory machines A, B and C manufacture respectively 25, 35 and 40% of the total. Out of their output 5, 4 and 2 Percent are defective bolts. A bolt is drawn from the produce and is found defective. What are the probabilities that is was manufactured by A, B and C ?*

Solution: Let E be the event that the bolt is defective and E_1, E_2, E_3 denote the events that the bolt is produced by A, B, C respectively.

We have $$P[E_1] = 0.25, P[E_2] = 0.35, P[E_3] = 0.40$$

$$P\left[\frac{E}{E_1}\right] = 0.05, P\left[\frac{E}{E_2}\right] = 0.04 \text{ and } P\left[\frac{E}{E_3}\right] = 0.02$$

It is required to find $P\left[\frac{E_1}{E}\right], P\left[\frac{E_2}{E}\right]$ and $P\left[\frac{E_3}{E}\right]$

By Baye's theorem, we have

$$P\left[\frac{E_1}{E}\right] = \frac{P[E_1]\,P\left[\frac{E}{E_1}\right]}{P[E_1]\,P\left[\frac{E}{E_1}\right] + P[E_2]\,P\left[\frac{E}{E_2}\right] + P[E_3]\,P\left[\frac{E}{E_3}\right]}$$

$$= \frac{(0.25)(0.05)}{(0.25)(0.05) + (0.35)(0.04) + (0.4)(0.02)}$$

$$\therefore \quad P\left[\frac{E_1}{E}\right] = \frac{125}{345} = \frac{25}{69}$$

Similarly,

$$P\left[\frac{E_2}{E}\right] = \frac{0.35 \times 0.04}{0.25 \times 0.05 \times 0.35 \times 0.04 + 0.40 \times 0.02} = \frac{140}{345} = \frac{28}{69}$$

and $$P\left[\frac{E_3}{E}\right] = 1 - \left\{P\left[\frac{E_1}{E}\right] + P\left[\frac{E_2}{E}\right]\right\} = 1 - \frac{25}{69} - \frac{28}{69} = \frac{16}{69}$$

Example 82: *The contents of urns I, II and III are as follow:*

1 white, 2 red and 3 black balls,

2 white, 3 red and 1 black ball, and

3 white, 1 red and 2 black balls.

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they come from urns I, II or III ?

Solution: Let E_1, E_2 and E_3 denote the events that the urns I, II and III are chosen, respectively and let E be the event that the two balls taken from the selected urn are white and red, then

$$P[E_1] = P[E_2] = P[E_3] = \frac{1}{3}$$

$$P\left[\frac{E}{E_2}\right] = \frac{{}^1C_1 \times {}^2C_1}{{}^6C_2} = \frac{2}{15}, P\left[\frac{E}{E_2}\right] = \frac{{}^2C_1 \times {}^3C_1}{{}^6C_2} = \frac{6}{15}$$

and $$P\left[\frac{E}{E_3}\right] = \frac{{}^3C_1 \times {}^1C_1}{{}^6C_2} = \frac{3}{15}$$

By Baye's theorem, we get

$$P\left[\frac{E_1}{E}\right] = \frac{P\,[E_1]\,P\left[\frac{E}{E_1}\right]}{\sum_{i=1}^{3} P\,[E_i]\,P\left[\frac{E}{E_i}\right]}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{2}{15}\right)}{\left(\frac{1}{3}\right)\left(\frac{2}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{6}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{15}\right)} = \frac{2}{11}$$

Similarly,

$$P\left[\frac{E_2}{E}\right] = \frac{\left(\frac{1}{3}\right)\left(\frac{6}{15}\right)}{\left(\frac{1}{3}\right)\left(\frac{2}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{6}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{15}\right)} = \frac{6}{11}$$

and $$P\left[\frac{E_3}{E}\right] = \frac{\left(\frac{1}{3}\right)\left(\frac{3}{15}\right)}{\left(\frac{1}{3}\right)\left(\frac{2}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{6}{15}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{15}\right)} = \frac{3}{11}$$

Problem Set

1. A coin is tossed 4 times. Define a sample space T which describes this experiment and give the subsets of T which correspond to the following events :

(i) More heads than tails are obtained.

(ii) Tails occur on the even numbered tosses.

(iii) Head occur on the third toss.

2. If $P[A]=0.9, P[B]=0.8$, show that $P[AB]\geq 0.7$.

[**Hint:** $P[AB]\geq P[A]+P[B]-1$ (Bon-ferroni inequality)].

3. Arrange the following in increasing order of magnitude :

$$P[A], P[A\cup B], P[A\cap B], P[A]+P[B]$$

4. For disjoint events A, B, C are the following permissible assignments of probabilities ?

$$P[A]=\frac{2}{3}, P[B]=\frac{1}{4}, P[C]=\frac{1}{6}$$

which is impossible.

[**Hint:** $P[A\cup B\cup C]=P[A]+P[B]+P[C]-P[A\cap B]-P[B\cap C]$

$$-P[A\cap C]+P[A\cap B\cap C]$$

$$=P[A]+P[B]+P[C]=\frac{2}{3}+\frac{1}{4}+\frac{1}{6}=\frac{13}{12}>1]$$

5. Given $P[A]=\frac{1}{3}, P[B]=\frac{3}{4}, P[A\cup B]=\frac{11}{12}$, find $P\left[\frac{A}{B}\right]$ and $P\left[\frac{B}{A}\right]$

$$\left[\textbf{Hint: } P[A\cup B]=P[A]+P[B]-P[AB]\Rightarrow P[AB]=\frac{1}{6}\right]$$

Now, use $P[AB]=\dfrac{P[A]\,P[B]}{A}=P[B]\,P\left[\dfrac{A}{E}\right]$.

6. An urn contains four tickets marked with numbers 112, 121, 211, 222 and one ticket is drawn at random. Let A_i $(i=1,2,3)$ be the event that its digit of the number of the ticket drawn is 1. Discuss the independence of the events A_1, A_2 and A_3.

$$\left[\textbf{Hint: } P[A_1]=\frac{2}{4}=\frac{1}{2}, P[A_2]=P[A_3]=\frac{1}{2}\right]$$

$A_1\cap A_2$ is the event that the first two digits in the number which the drawn ticket bears is each 1.

$\therefore$ $$P[A_1\cap A_2]=\frac{1}{4}=\frac{1}{2}\cdot\frac{1}{2}=P[A_1]\,P[A_2]\Rightarrow A_1, A_2$$

are independent. Similarly A_2, A_3, A_3, A_1 are independent.

Now, $P[A_1\cap A_2\cap A_3]=$ probability that all the three digits are each $1=0\neq P[A_1]\,P[A_2]\,P[A_3]$. Hence A_1, A_2, A_3 are not mutually independent.

7. There are two identical boxes containing respectively 4 white and 3 red balls, 3 white and 7 red balls. A box is chosen at random and a ball is drawn from it. Find the probability that the ball is white. If the ball is white. What is the probability that it is from first box?

[**Hint:** E_1 = event that it is first box $\Rightarrow P\,[E_1] = \frac{1}{2}$

E_2 = event that it is a second box $\Rightarrow P\,[E_2] = \frac{1}{2}$

E = event that the ball is white

$$P\left[\frac{E}{E_1}\right] = \frac{4}{7},\ P\left[\frac{E}{E_2}\right] = \frac{3}{10}\Bigg]$$

(i) $$P\,[E] = P\,[E_1]\,P\left[\frac{E}{E_1}\right] + P\,[E_2]\,P\left[\frac{E}{E_2}\right]$$

$$= \left(\frac{1}{2}\right)\left(\frac{4}{7}\right) + \left(\frac{1}{2}\right)\left(\frac{3}{10}\right) = \left(\frac{61}{140}\right)$$

(ii) By Baye's theorem, we have

$$P\left[\frac{E_1}{E}\right] = \frac{\left\{P\,[E_1]\,P\left[\frac{E}{E_1}\right]\right\}}{P\,[E]} = \frac{1}{2} \times \frac{4}{7} \times \frac{140}{61} = \frac{40}{61}$$

8. Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind. A colour blind person is chosen at random. What is the probability of his being male? (Assume male and female to be in equal numbers).

[**Hint:** $E_1 \equiv$ Person is a male, $E_2 \equiv$ Person is a female, $E \equiv$ Person is colour blind.

Then, $$P\,[E_1] = P\,[E_2] = \frac{1}{2},\ P\left[\frac{E}{E_1}\right] = 0.05,\ P\left[\frac{E}{E_2}\right] = 0.0025$$

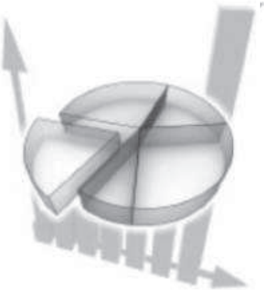
Find $P\left[\frac{E_1}{E}\right]$ by Baye's theorem.]

Answers

3.	$P\,[A \cap B],\ P\,[A],\ P\,[A] + P\,[B],\ P\,[A \cup B]$	**4.**	No	**5.**	$\frac{2}{9}, \frac{1}{2}$
6.	A_1, A_2, A_3 are not mutually independent.	**7.**	$\frac{61}{140}, \frac{40}{61}$	**8.**	0.95

❑❑❑

Chapter 8



Statistical Quality Control Methods

8.1 Quality

Quality is a measure of how closely an item conforms to standards. The standards may relate to time, materials, performance, reliability, appearance of the product.

As such in general, we may say that a product is considered to be good quality if it gives the desired services satisfactorily. The quality consisting the following aspects.

1. Customer's preference *i.e.* the degree or level to which it is preferred by the customers over similar products from other sources.
2. Degree of general excellence of the product as compared to other products of the same class.
3. Quality characteristic *i.e.* its appearance, life, taste, odour, reliability, etc.
4. Fitness for the purpose for which it is meant.

8.2 Method for the Measured the Quality

There are two methods to measure the quality of a manufactured product:

1. **The Numerical Measurement:** We decide about the quality of a product by making numerical measurements on it, such as length, height, radius, density, area, volume etc., and the items which are meeting with the specification are accepted.
2. **Attributal Measurement:** When any item is divided into two defective or effective classes then attraibutal measurement is applicable.

8.3 Quality Control

The Control of quality involves, measurement, feedback, comparison with standards and correctness when necessary.

Thus, the basic characteristic of quality control are:

1. To insure that the desired specifications of the product are achieved.
2. To Compare the product with standards and correct it when necessary.

8.4 Statistical Quality Control (S.Q.C.)

[B.C.A. (Lucknow) 2007, 2004; B.C.A. (Meerut) 2006, 2004, 2002;
B.C.A. (Kanpur) 2005; B.C.A. (Agra) 2003]

A quality control system performs inspection, testing and analysis to ensure that the quality of the products produced is as per the laid down quality standards. It is called "**Statistical Quality Control**" when statistical techniques are information system to satisfy the concept of prevention and control and improving upon product quality requires statistical thinking. In other words, the main purpose of statistical quality control is to device statistical methods for separating the variation of random nature from the other type, so that we may take appropriate action as quickly as possible when preventable causes are present in the system.

Statistical quality control (S.Q.C) is systematic as compared to guess-work of haphazard process inspection and the mathematical statistical approach neutralizes personal basis and uncovers poor judgement (S.Q.C) consists of three general activities.

1. Systematic collection and graphic recording of accurate data.
2. Analyzing the data
3. Practical engineering or management action, if the information obtained indicates significant deviations from the specified limits.

Modern techniques of statistical quality control and acceptance sampling having an important part to play in the improvement of quality, enhancement of productivity, creation of consumer confidence and development of industrial economy of the country.

The following statistical tools are generally used for the above purposes:

8.4.1 Frequency Distribution

Frequency distribution is a tabulation of the number of times a given quality characteristic occurs within the samples graphic representation of frequency distribution will show:

(i) Average quality

(ii) Spread of quality

(iii) Comparison with specific requirements

(iv) Process capability.

8.4.2 Control Chart

Control Chart is a graphical representation of quality characteristics, which indicates whether the process is under control or not.

8.4.3 Acceptance Sampling

The acceptance sampling is the process of evaluating a portion of the product or material in a lot for the purpose of accepting or rejecting the lot on the basis of conforming to a quality specification.

8.4.4 Analysis of Data

It includes analysis of tolerances, correlation, analysis of variance, analysis for engineering design, problem solving technique to estimate cause to troubles. Statistical methods can be used in arriving at proper specification limits of product, in designing the product, in purchase of raw-material, semi-finished and finished products, manufacturing processes, inspection, packaging, sales and also after sales service.

8.5 Advantages of Statistical Quality Control

[Kanpur B.C.A. 2004, 2005; B.C.A. (Lucknow) 2007, 2008; B.C.A. (Agra) 2006; B.C.A. (Meerut), 2001, 2003; B.C.A. (Rohtak) 2008, 2009, 2011]

1. **Efficiency:** The use of statistical quality control ensures rapid and efficient inspection at a minimum cost. It eliminates the need of 100% inspection of finished products because the acceptance sampling is statistical quality control exerts more effective pressure for quality improvement than it.
2. Increases out-put and reduces wasted machine and man hours.
3. It uncovers the cause of excessive variability in manufactured products fore costing trouble before rejections occur and reducing the amount of spoiled work.
4. Points out when and where 100% inspection, sorting or screening is required.
5. **Easy Detection of Faults:** In S.Q.C. after plotting the control charts ($\overline{X}, R, P, C, U$) etc., when the points fall above the upper control limits or below the lower control limit, it gives an indication of deterioration in quality, we can take immediately correct action.
6. Elimination of bottlenecks in the process of manufacturing.
7. Provides a common language that may be used by all three groups designers, production personnel and inspectors in arriving at a rational solution of mutual problems.
8. Adherance of Specifications.

8.6 Reasons for Variations in the Quality of a Product

Two extremely similar things are rarely obtain in nature. This fact hold good for production processes as well no production process is good enough to produce all items of products exactly alike. Variation in their quality is bound to occur. If we analyse this variation, we observe that there are certain causes responsible for this variation. The variations are due to two main reasons.

8.6.1 Chance or Random Cause or Non–Assignable Causes

The Variations due to chance causes are inevitable in any process or product. They are difficult to trace and to control also even under best conditions of production. This type of variation is unavoidable and no reason is possible to assign for such variation. We can say that this variation occurs in a random way. This type of variation can not be prevented and can not be detected and therefore is called allowable or tolerable and does not affect the quality and utility of the product.

If the variations are due to chance factors alone, the observations will follow a ‘‘normal curve’’ the knowledge of the behaviour of chance variation in the foundation on which control chart analysis rests. The conditions which produce these variations are accordingly said to be ‘‘**under control**’’.

If the variations in the data do not conform to a pattern that might reasonably by produced by chance causes, then in this case, conditions, producing the variations are said to be ‘‘out of control’’.

8.6.2 Assignable Cause or Preventable Causes

When the cause of variation in quality are such that they can be detected , assigned to some reasons and rectified to prevent further occurrence of such variability, they are known as assignable or preventable causes. The variations due to assignable cause possess greater magnitude as compare to those due to chance causes and can be easily traced or detected. The power of the shewhart control chart lies in its ability to separate out these assignable causes of quality variations *e.g.* in length thickness, weight or diameter of a component.

The Variations due to assignable causes may be because of following factors:

(i) Differences among machines

(ii) Differences among workers

(iii) Differences in their relationship to one another. There variations may also be caused due to change in working conditions, mistake on the part of operator etc. Thus affects the quality and utility of the product and cause worry to the manufacture.

8.7 Techniques of Statistical Quality Control

[B.C.A. (Meerut) 2003, 2005; B.C.A. (Lucknow) 2004, 2008]

In statistical quality control our aim is to bring out the assignable causes of variation as soon as they occur, so that the actual variation may be supposed to be due to chance causes alone.

To control the quality characteristics of the product, there are two main techniques.

8.7.1 Process Control

It is the process of monitoring and measuring variability in the performance of a process or a machine through the interpretation of statistical techniques and it is employed to manage in-process quality. This technique ensures the production of requisite standard product and make use of control charts. In this type of problem our aim is to control the manufacturing process so that the proportion of defective items is not large.

8.7.2 Product Control

This technique is concerned with inspection of already produced goods to ascertain whether they are fit to be despatched or not. To achieve the objectives, it makes use of sampling inspection plans. In this type of problem, we like to ensure that lots manufactured goods do not contain an excessively large proportion of defective items, Process control is achieved mainly through the technique of control charts, whereas product control is achieved through sampling plans. This process can be understand by following chart:

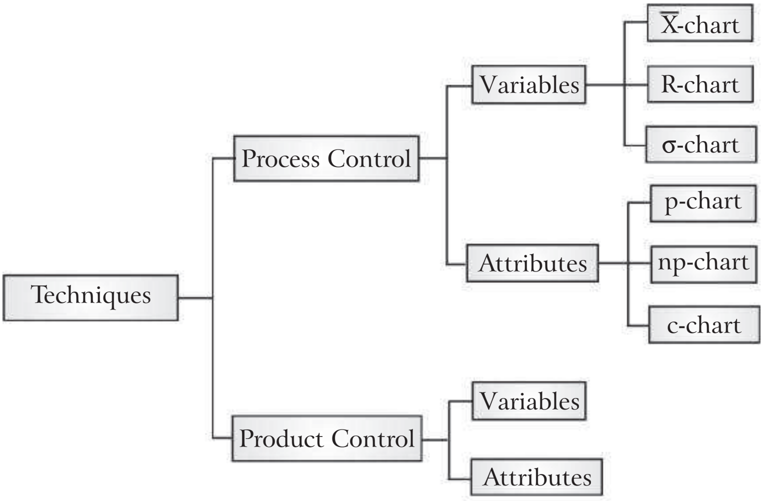


Fig. 1

8.7.3 Control Chart

A control chart is a graphical representation of the collected information that remove the variation in processing and warns if there is any departure from the specified tolerance limits. We can say control chart is a device for attaining quality control or is a device to judge whether the statistical control has been attained.

8.7.4 Control Limits

The control limits on the chart are so placed as to disclose the presence or absence of the assignable causes of quality variation which make the diagenosis possible and brings substantial improvements in the product quality and reduction of spailage and rework. Control chart tells when to leave the process alone and thus prevents unnessarily frequent adjustments that tend to increase the variability of the process rather than to decrease it. A control chart is simply a frequency distribution of the observed value plotted as points in order of occurrence so that each value has its own identity relative to the time of observation. The points on the control charts may or may not be connected.

In control charts, the quality characteristic (Q) of manufactured product follows a distribution only when random causes are present in the process. This distribution has the mean $\overline{Q}$ and standard deviation σ_Q. If this distribution is normal the limits $\overline{Q} \pm 3\sigma_Q$ consists of 99.73 points of the distribution. The limits $\overline{Q} + 2\sigma_Q$ and $\overline{Q} + \sigma_Q$ consists of 95.44% and 68.26% points of the distribution respectively. If the distribution of Q is not normal, the limits $\overline{Q} \pm 3\sigma_Q$ consists 88.89% points of the distribution. Therefore, in every case, these limits contain all the points of distribution and if any point is beyond these limits, we find that not only random causes are present in the process but the assignable causes are also present. In other words, we can say that the process is out of control. If there is no point outside the limit, then the process is under control. From above, we conclude that in SQC we don't control the quality but we control the process.

By considering these data, we state earlier a central line $\overline{Q}$, upper control limit $\overline{Q} + 3\sigma_Q$ and lower control limit $\overline{Q} - 3\sigma_Q$ are drawn as horizontal straight lines on a graph paper and the number of samples known as homogenous sub-groups taken from different shifts of the process are plotted on the same graph paper. The sub-group number is taken on X-axis and quality characteristic Q is taken on Y-axis. The graph, therefore obtained is known as control chart. An example of control chart is shown on the next page.

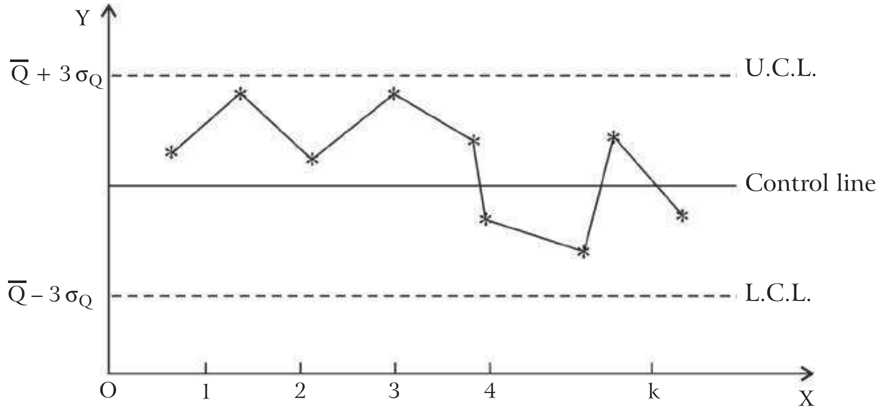


Fig. 2: Sub-group number

There are many types of control charts designed for different control situations. Most commonly used control charts are:

(i) **Control Charts for Variables:** They are useful to measure quality characteristics and to control automatic process. It includes $\overline{X}$ and R-charts and charts for $\overline{X}$ and σ.

(ii) **Control Charts for Attributes:** It includes P-chart for fraction defective. A fraction defective control chart discloses erratic fluctuations in the quality of inspection which may result in improvement in inspection practice and inspection standards.

It also includes C-chart for number of defects per unit.

8.8 Object of Control Charts

1. $\overline{X}$ and R or $\overline{X}$ and σ charts are used in combination of control process. $\overline{X}$-chart shows the variation in the average of samples.

   It is the most commonly used variables chart. R-chart shows the uniformity or consistency of the process. It is the chart for measure of spread. σ-chart shows the variation of process.

2. To secure information to be used in establishing or changing production procedures.

3. To secure information to be used in establishing or changing inspection procedure or acceptance procedure or both.

4. To determine whether a given process can meet the existing specifications without a fundamental change in the production line or to tell whether the process is in control and it so, at what dispersion.

8.9 Type of Control Charts

Control Charts is divide into two groups:

1. **Control Charts for Variables**

   When quality is measured through numerical measurement, control charts are called control charts for variables. For example measurements of resistance of wire, density of liquid, volume of solid body are of continuous type and follow the normal probability law. For quality control of such data, the following control charts are used:

   (i) Control chart for mean ($\overline{X}$)

   (ii) Control chart for S.D. (σ-chart)

   (iii) Control chart for range (R-chart)

2. **Control Charts for Attributes**

   If quality characteristic is on attribute, and each item is regarded as either defective or non-defective.

   There are three types of control charts in practical.

   (i) Control charts for function defective (p-chart)

   (ii) Control charts for number of defective (np-chart)

   (iii) Control charts for number of defective (C-chart)

8.9.1 Construction of Control Charts for Variables

First of all, a random sample of size n is taken during a manufacturing process over a period of time and quality measurements $x_1, x_2 \ldots, x_n$ are noted.

$$\text{Sample mean} = \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\text{Sample range} = R = x_{\max} - x_{\min}$$

If the process is found stable, k consecutive samples are selected and for each sample, $\bar{x}$ and R are calculated.

Then find mean ($\bar{x}$) and sample range (R) as follow:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_k}{k} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$$

and $$\overline{\boldsymbol{R}} = \frac{R_1 + R_2 + \ldots + R_h}{h} = \frac{1}{h}\sum_{i=1}^{h} R_i$$

This $\bar{\bar{x}}$ is called control line (C.L.)

For $\overline{X}$-chart

$$\text{Control line} = \begin{cases} \bar{x}, \text{when tolerance limits are not given} \\ \mu, \text{when toterence limit are given} \end{cases}$$

where $$\mu = \frac{1}{2}\,[\text{L.C.L} + \text{U.C.L}]$$

$$\text{L.C.L.} = \text{lower control limit}$$

$$\text{U.C.L} = \text{upper control limit}$$

Now, L.C.L. (for $\overline{X}$ -chart) $= \bar{\bar{x}} - A_2\ \overline{R}$

and U.C.L. (for X-chart) $= \bar{\bar{x}} + A_2\ \overline{R}$

A_2, depends on sample size n and can be found from the following table:

Sample Size (n)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A_2	1.88	1.02	0.73	0.58	0.48	0.42	0.37	0.34	0.31	0.29	0.27	0.25	0.24	0.22	0.21	0.20	0.19	0.19	0.18

For R-chart: Control line (C.L.) $= R$

and L.C.L. (for R-chart) $= D_3\overline{R}$

U.C.L. (for R-chart) $= D_4\overline{R}$

Where D_3 and D_4 depend on sample size are found from the following table:

Sample Size (n)	D_3	D_4	d_2
2	0	3.27	1.13
3	0	2.57	1.69
4	0	2.28	2.06
5	0	2.11	2.33
6	0	2.00	2.53
7	0.08	1.92	2.70
8	0.14	1.86	2.85
9	0.18	1.82	2.97
10	0.22	1.78	3.08
11	0.26	1.74	3.17
12	0.28	1.72	3.26
13	0.31	1.69	3.34
14	0.33	1.67	3.41
15	0.35	1.65	3.47
16	0.36	1.64	3.53
17	0.38	1.62	3.59
18	0.39	1.61	3.64
19	0.40	1.60	3.69
20	0.41	1.59	3.74

To Compute upper and lower process tolerance limits for the value of x, we find

$$\text{L.T.L} = \bar{\bar{x}} - \frac{3\bar{R}}{d_2}, \qquad \text{U.T.L} = \bar{\bar{x}} + \frac{3\bar{R}}{d_2}$$

Where d_2 is found from the above table.

The process capability is given by $6\sigma = 6\frac{\bar{R}}{d_2}$.

Where σ is standard deviation. While plotting the $\bar{X}$-chart the control line on the $\bar{X}$-chart should be drawn as a solid horizontal line at $\bar{X}$. The upper and lower control limits for $\bar{X}$-chart should be drawn as dotted horizontal lines at the computed values.

Similarly, for R-chart, the control line should be drawn as a solid horizontal line at $\bar{R}$. The upper control limit should be drawn as dotted horizontal line at the computed value of UCL_R.

If sub-group size is 7 or more, the lower control limit should be drawn as dotted horizontal line at LCL_R. However, if the sub-group size is ≤ 6, the lower control limit for R is zero.

Conclusion: If the observed values fall within the control limits, the process is assumed to be in control at that moment of production. If the observed values fall out side the control limits, the process is said to be out of control at that moment and presence of assignable causes is indicated in the process. However, points which fall outside the control limits do not necessarily represent rejected material but only signas that some corrective action is required to prevent manufacturing faculty parts.

8.10 Control Charts for Attributes

[B.C.A. (Meerut) 2005, 2003]

Following control charts will be discussed here:

1. P-chart
2. np-chart
3. C-chart
4. u-chart

As an alternative to $\bar{X}$ and R-chart and as a substitute when characteristc is measured only by attribute, a control chart based on fraction defective p is used, called P-chart.

$$p = \frac{\text{Number of defective articles found in any inspection}}{\text{Total number of articles actually inspected}}$$

1. Control Limits (3σ limits) on *p*–Chart

We know that for binomial distribution, the mean value of total number of defectives in a sample *n is np* and standard deviation is $\sqrt{npq}$.

where $p+q=1$ or $q=1-p.$

$\because$ Mean value of fraction defective is p and S.D.

$$\sigma_p = \frac{\sqrt{np(1-p)}}{n} = \frac{\sqrt{p(1-p)}}{n}$$

$$\therefore \qquad \text{C.L.} = p$$

The upper and lower limits for P-chart are:

$$\text{U.C.L.}_p = p + 3\sigma_p = p + 3\sqrt{\frac{p(1-p)}{n}}$$

and $$\text{L.C.L}_p = p - 3\,\sigma_p = p - 3\sqrt{\frac{p(1-p)}{n}}$$

2. Control Limits for *np*–Chart

Whenever sub-group size is variable, P-chart is used but if it is constant the chart for actual number defectives called *np* called *np* chart is used

$$\text{C.L.} = n\overline{p} \quad \text{where} \quad \overline{p} = \frac{\Sigma np}{\Sigma n}$$

$$\text{U.C.L.}_{np} = n\overline{p} + 3\sigma_{np} = n\overline{p} + 3\sqrt{n\overline{p}\,(1-\overline{p})}$$

where $$\sigma_{np} = n\sigma_p$$

and $$\text{L.C.L.}_{np} = n\overline{p} - 3\sqrt{n\overline{p}\,(1-\overline{p})}$$

3. Control Limits for C–Chart

(i) Difference between a defect and defective.

(ii) An item is called defective if it fails to conform to the specifications in any of the characteristics. Each characteristic that does not meet the specifications is a defect. An item is defective if it contains atleast one defect. The *np* chart applies to the number of defectives in sub-groups of constant size while C-chart applies to the number of defects in a sub-group of constant size.

(iii) **Basis for control limits on C-chart:** Control limits on C-chart are based on **Poisson distribution**.

(iv) **Calculation of control limits on C-chart:** Standard deviation $\sigma_c = \sqrt{\overline{C}}$

Thus 3σ limits on a C-chart are:

$$\text{U.C.L.}_C = \overline{C} + 3\sqrt{\overline{C}}$$

$$\text{L.C.L.}_C = \overline{C} - 3\sqrt{\overline{C}}$$

and control line C.L. $= \overline{C}$

where $\overline{C} = \dfrac{\text{number of defects in all samples}}{\text{total number of samples}}$

4. **u-chart:** When the sub-group size varies from sample to sample, it is necessary to use u-charts. The control limits on u-chart will however vary. If C is total number of defects found in any sample and n is number of inspection units in a sample.

$$\overline{u} = \frac{c}{n} = \frac{\text{number of defective in a sample}}{\text{number of units in a sample}}$$

The control limits on u-chart are

$$\text{U.C.L.}_u = \overline{u} + 3\sqrt{\frac{\overline{u}}{n}}; \quad \text{L.C.L}_\text{u} = \overline{u} - 3\sqrt{\frac{\overline{u}}{n}}$$

and control line C.L. $= \overline{u}$

There are two ways of obtaining these limits. One way is to use the value of ranges and another way is to use the values of standard deviations.

On the Basis of R (Range)	On the Basis of s (S.D)
From the sampling distribution of range $E(R) = d_2\sigma_p$ Where d_2 is a constant depending upon the sample size. Thus an estimate of σ_p can be found from $\overline{R}$ by $\sigma_p = \overline{R}/d_2$ Also $\overline{\overline{x}}$ gives an unbiased estimate of the population mean μ_p. $\therefore \quad E(\overline{\overline{x}}) = \dfrac{\sum_{i=1}^{k} E(\overline{x}_{2i})}{k} = \mu_p$	If the control limits are to be found on the basis of $\overline{s}$ except $\overline{R}$, then the estimate of σ_p can be found from $\overline{s} = c_2\,\sigma_p$ $\Rightarrow \quad \sigma_p = \dfrac{\overline{s}}{c_2}$ where $C_2 = \sqrt{\dfrac{2}{n}} = \dfrac{\left(\dfrac{n-2}{2}\right)!}{\left(\dfrac{n-3}{2}\right)!}$

Solved Examples

Example 1: *Adriling machine bores hales with a mean diameter* 0.5230 *cm and a standard deviation of* 0.0032 *cm. calculate the* 2*-sigma and* 3*-sigma upper and lower control limits for means of sample of* 4.

[B.C.A. (Lucknow) 2008, 2006; B.C.A. (Rohtak) 2008;

B.C.A. (Kanpur) 2006; B.C.A. (Meerut) 2005; B.C.A. (Agra) 2004]

Solution: Mean diameter $\bar{\bar{x}} = 0.5230$ cm

$$\text{S.D.}\,(\sigma) = 0.0032 \text{ cm}$$

$$n = 4$$

(i) 2-sigma limits are:

$$\text{C.L.} = \bar{\bar{x}} = 0.5230 \text{ cm}$$

$$\text{U.C.L.} = \bar{\bar{x}} + 2\frac{\sigma}{\sqrt{n}} = 0.5230 + 2 \times \frac{0.0032}{\sqrt{4}} = 0.5262 \text{ cm}$$

$$\text{L.C.L.} = \bar{\bar{x}} - 2\frac{\sigma}{\sqrt{n}} = 0.5230 - 2 \times \frac{0.0032}{\sqrt{4}} = 0.5198 \text{ cm}$$

(ii) 3-sigma limits are:

$$\text{C.L.} = \bar{\bar{x}} = 0.5230 \text{ cm}$$

$$\text{U.C.L.} = \bar{\bar{x}} + 3\frac{\sigma}{\sqrt{n}} = 0.5230 + 3 \times \frac{0.0032}{\sqrt{4}} = 0.5278 \text{ cm}$$

$$\text{L.C.L.} = \bar{\bar{x}} - 3\frac{\sigma}{\sqrt{n}} = 0.5230 - 3 \times \frac{0.0032}{\sqrt{4}} = 0.5182 \text{ cm}$$

Example 2: *In the production of certain rods, a process is said to be in control is the outside diameters have a mean* 2.5" *and a s.d. of* 0.002".

(i) *Find the control limits for the mean of random samples of size* 4.

(ii) *Mean of* 10 *random samples taken at regular interval were:* 2.5014, 2.5022, 2.4995, 2.5076, 2.504, 2.4993, 2.4962, 2.4966 *and* 2.4971, *was the process even out of control.*

Solution: Given that $\bar{x} = 2.5'', \sigma = 0.002'', n = 4$

(i) Control limits for $\overline{X}$-chart

$$\text{The control line (C.L.)} = \bar{\bar{x}} + \frac{3\sigma}{\sqrt{n}} = 2.5 + \frac{3 \times 0.002}{\sqrt{4}}$$

$$= 2.5 + 0.003 = 2.503''$$

$$\text{L.C.L.} = \bar{\bar{x}} - \frac{3\sigma}{\sqrt{4}} = 2.5 - \frac{3 \times 0.002}{\sqrt{4}} = 2.5 - 0.0003 = 2.497''$$

(ii) On the basis of control line U.C.L and L.C.L. and plotting the value of 10 means, it is found that process is out of control corresponding to sample number 4, 5, 8 and 9

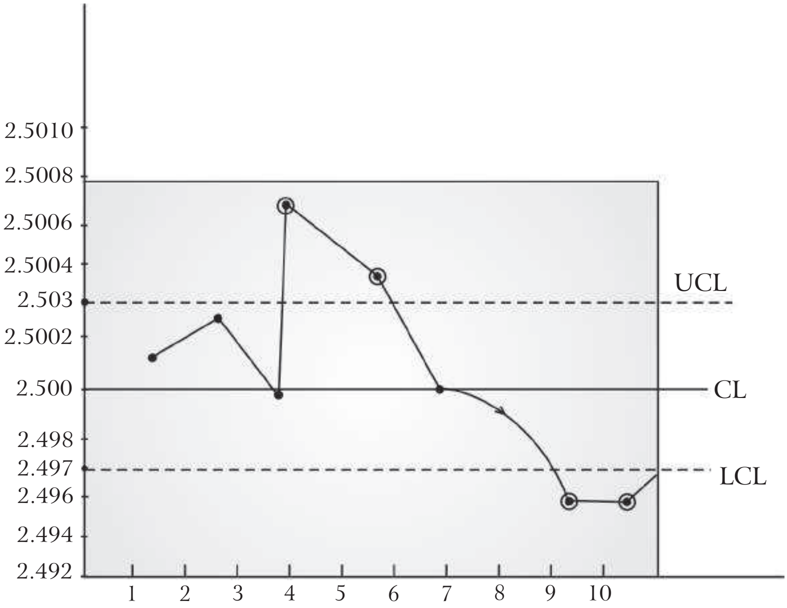


Fig. 3: Sample number

Mean chart showing process out of control.

Example 3: *The following are the mean lengths and ranges of lengths of a finished product from 10 samples each of size 5. The specification limits for length are 200 ± 5 cm construct $\overline{X}$ and R-chart and examine whether the process is under control and state your recommendations.*

Sample No. 1	1	2	3	4	5	6	7	8	9	10
Mean ($\overline{X}$)	201	198	202	200	203	204	199	196	199	201
Range ($\overline{R}$)	5	0	7	3	3	7	2	8	5	6

Assume for $n = 5$, $A_2 = 0.58$, $D_4 = 2.11$ *and* $D_3 = 0$

[B.C.A. (Lucknow) 2010, 2006, 2002]

Solution:

(i) Control limits for $\overline{X}$ chart:

Control limits (C.L.) = 200

∵ Tolerance or specification limits are given:

$$\therefore \qquad \mu = 200$$

$$\text{U.C.L.}_{\bar{x}} = \bar{\bar{x}} + A_2 \overline{R} = \mu + A_2 \overline{R}$$

$$\text{L.C.L.}_x = \frac{R_1 + R_2 + \ldots + R_{10}}{10} = \frac{46}{10} = 4.6$$

Then, $$\text{U.C.L.}_{\bar{x}} = 200 + (0.58 \times 4.6) = 202.668$$

$$\text{L.C.L.}_{\bar{x}} = 200 - (0.58 \times 4.6) = 197.332$$

(ii) Control limits for *R*-chart:

$$\text{Control limit (C.L.)} = \overline{R} = 4.6$$

$$\text{U.C.L.}_R = D_4\overline{R} = 2.11 \times 4.6 = 9.706$$

$$\text{L.C.L.}_R = D_3\overline{R} = 0 \times 4.6 = 0$$

The $\overline{X}$ and R-charts are drawn below:

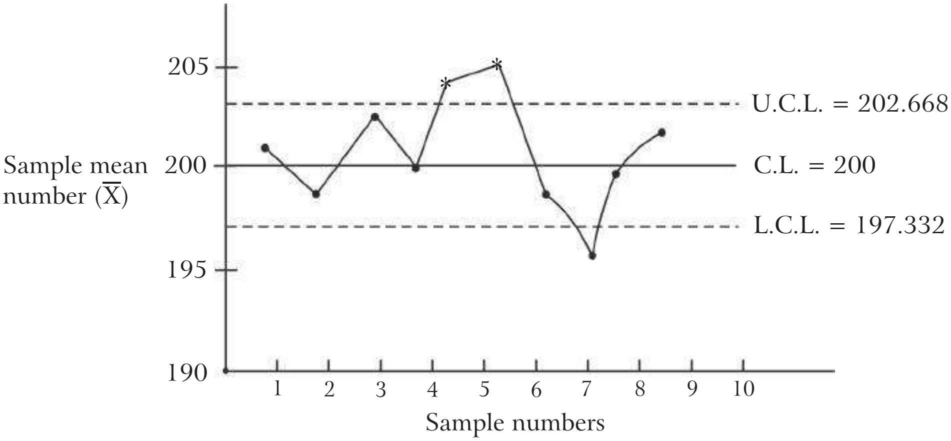


Fig. 4: X-chart

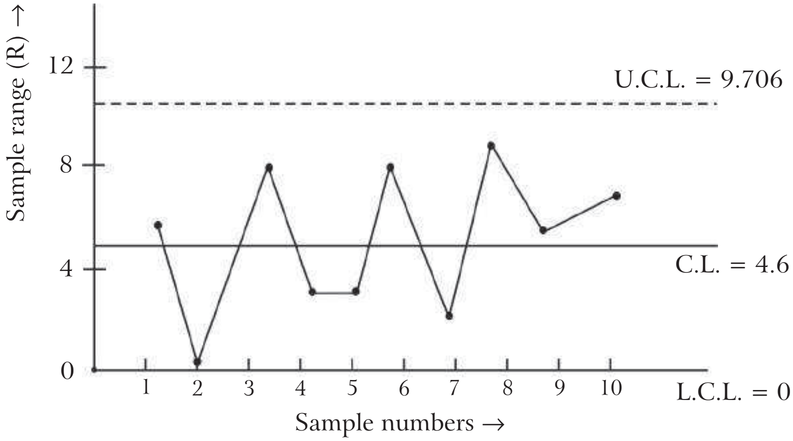


Fig. 5: R-chart

Example 4: *Construct a control chart for mean and range for the following data on the basis of fuses, sample of 5 being taken every hour (each set of 5 has been arranged in ascending order of magnitude). Comment on whether the production seems to be under control. Assuming that these are the first data:*

42	42	19	36	42	71	60	18	15	69	64	61
65	45	24	54	51	74	60	20	30	109	90	78
75	68	80	69	57	75	72	27	39	113	93	94
78	72	81	77	59	78	95	42	62	118	109	109
87	90	81	84	78	132	138	60	84	153	112	136

Solution:

Sample No.	Sample Observations					Total	Sample Mean $(\overline{X})$	Sample Range $(\overline{R})$
1	42	65	75	78	87	347	69.4	45
2	42	45	68	72	90	317	63.4	48
3	19	24	80	81	81	285	57.0	62
4	36	54	69	77	84	320	64.0	48
5	42	51	57	69	78	297	59.4	36
6	71	74	75	78	132	430	86.0	61
7	60	60	72	95	138	425	85	78
8	18	20	27	42	60	167	33.4	42
9	15	30	39	62	84	230	46	69
10	69	109	113	118	153	562	112.4	84
11	64	90	93	109	112	468	93.6	48
12	61	78	94	109	136	478	95.6	75
Total						**865.2**	**696**	

From the above data we obtain

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} X_i}{k} = \frac{865.2}{12} = 72.1$$

$$\overline{R} = \frac{\sum_{i=1}^{k} R_i}{k} = \frac{696}{12} = 58$$

Control limits for $\overline{X}$-Chart

$$\text{U.C.L.} = \overline{\overline{X}} + A_2\,\overline{R} = 72.1 + 0.58 \times 58 = 105.74$$

where $A_2 = 0.58$ at $n = 5$

$$\text{L.C.L.} = \overline{\overline{X}} - A_2\,\overline{R} = 72.1 - 0.58 \times 58 = 38.46$$

$$\text{C.L} = \overline{\overline{X}} - 72.1$$

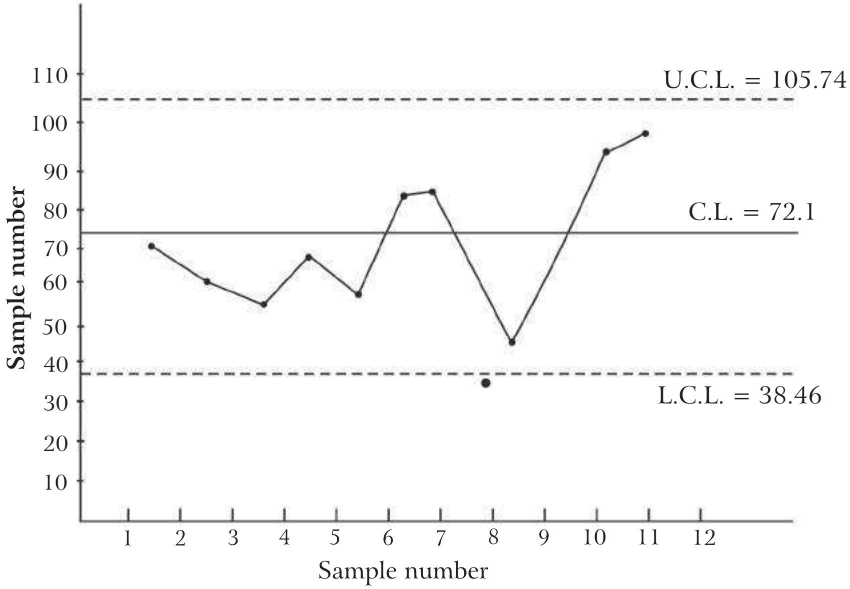


Fig. 6

For *R*-chart

From the about chart it is clear that process is out of control corresponding to the 8th and 10th samples.

$$\text{U.C.L.} = D_4\,\overline{R} = 2.11 \times 58 = 122.38$$

(The value of D_4 for $n = 5$ is 2.11)

$$\text{L.C.L.} = D_3\,\overline{R} = 0 \times 59.67 = 0$$

$$\text{C.L.} = 58$$

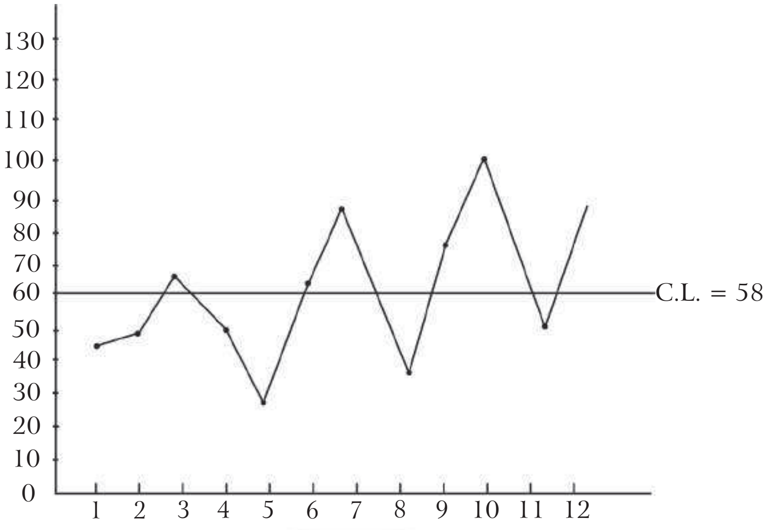


Fig. 7

Since all the sample points fall within the control limits the process is within control.

Example 5: *20 random samples of 5 units drawn from each lot of wire gave the following mean diameters and ranges. Find out control limits for* $\overline{X}$ *and R-chart and comment.*

Sub-group	$\overline{X}$	*R*	*Sub-group*	$\overline{X}$	*R*
1	61.2	10	11	60.8	13
2	60.8	9	12	65.2	10
3	61.8	5	13	67.8	10
4	62.2	9	14	66.0	9
5	59.2	11	15	64.2	6
6	62.0	14	16	64.8	8
7	61.6	12	17	69.6	13
8	66.0	6	18	59.4	11
9	62.0	15	19	61.0	15
10	66.0	16	20	61.8	13

Solution: We have 20 samples and size of each is 4

i.e., $k = 20, n = 4, \Sigma\overline{X} = 1263.4, \ \Sigma R = 215$

$\therefore$ $\overline{\overline{X}} = \frac{\Sigma\overline{X}}{k} = \frac{1263.4}{20} = 63.17$

and $\overline{R} = \frac{\Sigma R}{k} = \frac{215}{20} = 10.75$

Control limits for $\overline{X}$-chart

$$\text{Control line } (\overline{\overline{X}}) = 63.17$$

$$\text{U.C.L.} = \overline{\overline{X}} + A_2 R = 63.17 + 0.58 \times 10.75$$

$$(\text{where } n = 5, \text{ and } A_2 = 0.58)$$

$$\because \quad \text{U.C.L.} = 63.17 + 6.235 = 69.5$$

$$\text{L.C.L.} = \overline{\overline{X}} - A_2 \overline{R} = 63.17 - 0.58 \times 10.75$$

$$= 63.17 - 6.235$$

$$= 56.93$$

Where $n = 5$ and $D_3 = 0$ from table.

From the value of $\overline{X}$ given above, it is found that the process is out of control corresponding to sample number 17 and from the value of R given above, it is found that the process within the control.

Example 6: *In a blade manufacturing factory,* 1,000 *blades are examined daily. Draw the np-chart for the following table and examine whether the process is under control ?*

Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of Defective Blades	9	10	12	8	7	15	10	12	10	8	7	13	14	15	16

[B.C.A. (Kurukshetra) 2011, 2007; B.C.A. (Lucknow) 2008, 2006, 2005;
B.C.A. (Kanpur) 2004; B.C.A. (Meerut) 2004, 2002]

Solution: We have $n = 1{,}000$

$$\Sigma np = \text{total number of defectives} = 166$$

$$\Sigma n = \text{total number inspected} = 1{,}000 \times 15$$

$$\therefore \quad \overline{p} = \frac{\Sigma np}{\Sigma n} = \frac{166}{1{,}000 \times 15} = 0.011$$

$$\therefore \quad n\overline{p} = 1{,}000 \times 0.011 = 11$$

Cotrol limits are $\text{C.L.} = n\overline{p} = 11$

$$\text{U.C.L.}_{np} = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} = 11 + 3\sqrt{11(1-0.011)} = 20.894$$

$$\text{L.C.L.}_{np} = n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} = 11 - 3\sqrt{11(1-0.011)} = 1.106$$

The *np*-chart is drawn in the figure and we see that all the points lie within the control limits and process is under control.

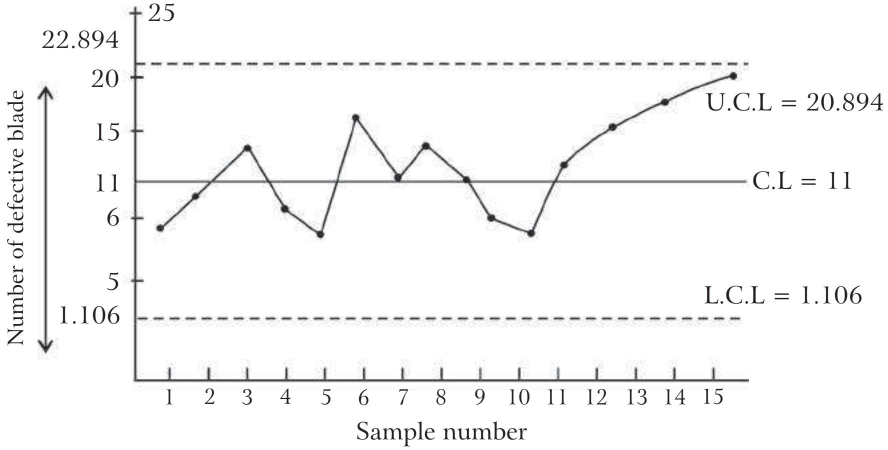


Fig. 8: *np*-chart

Example 7: *If the average function defective of a large sample of a product is* 0.1537, *calculate the control limits given that sub-group size is* 2,000.

[B.C.A. (A.M.U.) 2011, 2007, 2005; B.C.A. (Meerut) 2008, 2004]

Solution: Average fraction defective $\overline{p} = 0.1537$

sub-group size is 2,000 *i.e.* $n = 2{,}000$

control line $\text{C.L.} = n\overline{p} = 2{,}000 \times 0.1537 = 307.4$

$$\text{U.C.L.}_{np} = n\overline{p} + 3\sigma_{np} = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})}$$

$$= 307.4 + 3\sqrt{307.4(1-0.1537)} = 307.4 + 48.38774204$$

$$= 355.787742$$

and $\text{L.C.L.}_{np} = n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} = 307.4 - 48.38774204$

$$= 259.012258$$

Example 8: *A control chart for defects per unit 'u' uses probability limits corresponding to probabilities of* 0.975 *and* 0.025. *The control line on the control chart is at* $\overline{u} = 2.0$. *The limits vary with the the value of n. Determine the correct position of these upper and lower control limits when* $n = 5$ *(assume* $\sigma = 1.96$*)*

Solution: $\text{U.C.L}_u = \overline{u} + \sigma\sqrt{\dfrac{\overline{u}}{n}} = 2 + 1.96\sqrt{\dfrac{2}{5}} = 3.299$

and $\text{L.C.L}_u = 2 - 1.96\sqrt{\dfrac{2}{5}} = 0.761$

Example 9: *In a manufacturing process, the number of detectives found in the inspection of 20 lots of 100 samples is given below:*

Lot No.	*No. of Defectives*	*Lot. No.*	*No. of Defectives*
1	5	11	7
2	4	12	6
3	3	13	3
4	5	14	5
5	4	15	4
6	6	16	2
7	9	17	8
8	15	18	7
9	11	19	6
10	6	20	4

(i) Determine the control limits of p-chart and state whether the process is in control.

(ii) Determine the new value of mean fraction defective if some points are out of control. Computer the corresponding control limits and state whether the process is still in control or not.

(iii) Determine the sample size when a quality limit not worse than 9% is desirable and a 10% bad product will not be permitted more then three times in thousand.

Solution:

(i) We have $\bar{p} = \dfrac{\text{total number of defectives}}{\text{total number of items inspected}} = \dfrac{120}{20 \times 100} = 0.06$

$$\text{U.C.L.}_{p} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.06 + 3\sqrt{\frac{0.06(1-0.06)}{100}}, (n = 100)$$

$$= 0.13095$$

$$\text{L.C.L.}_{p} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.06 - 3\sqrt{\frac{0.06(1-0.06)}{100}} = -0.01095$$

Since the defective can not be negative then $\text{L.C.L.}_{p} = 0$ after observing the values of defectives in the given examples, it is clear that only 8th lot having fraction defective $\dfrac{15}{100} = 0.15$ will go above U.C.L._{p}

(ii) After eliminating the 8th lot revised value for $\bar{p} = \dfrac{120 - 15}{100 \times 19} = 0.056$

The revised control limits will be

$$\text{U.C.L.}_{p} = 0.056 + 3\sqrt{\frac{0.056(1-0.056)}{100}} = 0.125$$

$$\text{L.C.L}_p = 0.056 - 3\sqrt{\frac{0.056\,(1-0.056)}{100}} = -0.013 = 0$$

It is clear that all the points are within control limits

$\therefore$ revised quality level $\overline{p} = 0.056$

(iii) Since a probability that a defective worse than 9% defective quality will not be permitted is more than 3 times in thousand (0.3%) is corresponding 36 limits

$$\overline{p} + 3p = 0.09$$

or $$0.056 + 3\sqrt{\frac{0.056\,(1-0.056)}{n}} = 0.09$$

or $$\sqrt{\frac{0.056 \times 0.944}{n}} = \frac{0.034}{3}$$

Squaring $$\frac{0.056 \times 0.944}{n} = \left(\frac{0.034}{3}\right)^2 = (0.01133)^2$$

$$\Rightarrow \qquad n = \frac{0.056 \times 0.944}{0.01133 \times 0.01133} = 333.$$

8.11 Control Charts for Variables Versus Control Charts for Attributes

Variable Charts	Attribute Chart
1. Separate charts are constructed for each quality characteristics.	1. A single chart is considered sufficient to serve the purpose.
2. The data is collected purposively for the use of variable charts.	2. Data collected for other purposes may also be used.
3. Cost of collecting information on units is large.	3. Cost of collecting information on units is very small.
4. As the control charts are sensitive to assignable causes, these charts provides better quality control.	4. These charts are less sensitive to assignable causes, these charts are not much effective in quality control.
5. Usually a sample of size 4 or 5 serve the purpose.	5. Usually a large sample size works for correct conclusions.
6. In general, sub-group size is taken to be equal.	6. The need for equal size of samples is not much required as in case of variables.
7. Computational procedure is lengthy and time consuming.	7. Computational procedure is easy and simple.

8.12 Advantage of Statistical Quality Control

We know that by S.Q.C, the quality of a certain product is improved by bringing the process under control. The following are some advantages of S.Q.C.

1. By correcting (repairing) the plaint at an appropriate time, a considerable amount of cost per unit is reduced.
2. A considerable amount of excess raw-material is avoided.
3. The inspection of each and every unit is not practicable and reduce the efficiency of the quality inspectors. S.Q.C. keeps the workers alert.
4. S.Q.C. beeps continuous vigilance on the quality of the product. As soon as the presence of assignable causes is feit the concerned engineer is informed and the related part is corrected. In this way, the quality is improved and probability of the company increases.
5. The number of defective items become less and amount of scrap reduces.
6. S.Q.C. is economical in the sense that only a fraction of items produced is inspected.
7. It provides protection against losses to the producer as well as to consumer.

8.13 Limitation of Statistical Quality Control

A number of advantages are discussed above and the usefulness of S.Q.C. is focused. But S.Q.C. is not only the single technique for controlling all type of evils in production. A part from S.Q.C. many more consideration are made in the production of units. We can say that S.Q.C. is one of the many considerations which are made in the production system. That is S.Q.C. is a part of the production process rather than the production being a pat of S.Q.C.

Control limits for $\overline{X}$-chart

On the Basis of $\overline{R}$	On the Basis of $\bar{s}$
U.C.L. $= \bar{\bar{x}} + \frac{\overline{R}}{d_2} = \bar{\bar{x}} + A_2\overline{R}$	U.C.L. $= \bar{\bar{x}} + \left(\frac{3}{\sqrt{n}c_2}\right)\bar{s} = \bar{\bar{x}} + A_1\bar{s}$
L.C.L. $= \bar{x} - \frac{\overline{R}}{d_2} = \bar{\bar{x}} - A_2\overline{R}$	L.C.L. $= \bar{\bar{x}} - \left(\frac{3}{\sqrt{n}c_2}\right)\bar{s} = \bar{x} - A_2\bar{s}$
C.L. $= \bar{\bar{x}}$ or $\mu_p = \bar{\bar{x}}$	C.L. $= \bar{\bar{x}}$

After obtaining the upper and lower control limits and all the values of statistics, sub-groups (samples) numbers are taken on x-axis and the value of statistics on y-axis.

All the points thus plotted are connected and chart is found. If all the points lie within the control limits, we can say that the process is under control and if any point goes outside the control limits, we can say that the process is out of control.

And some assignable causes are present in the system that must be corrected.

8.14 Control Chart for Number of Defects per Unit (C-Chart)

If c_i is the number of defects observed in i-th unit, then then estimate of λ is given by $\bar{c}$ as it is an unbiased estimate of λ *i.e.* E(c),= λ.

The value of $\bar{c}$ is calculated by the following formula:

$$\bar{c} = \frac{\text{total number of defects}}{\text{number of units observed}} = \frac{c_1 + c_2 + \ldots + c_k}{k}$$

Thus, 3-sigma are given by:

$$\text{U.C.L.} = \bar{c} + 3\sqrt{\bar{c}}$$

$$\text{L.C.L.} = \bar{c} - 3\sqrt{\bar{c}}$$

$$\text{Control line} = \bar{c}$$

If L.C.L. comes out to be negative, it is considered as zero as the number of defects can not be negative.

Example 10: *The number of defects in* 24 *rolls of cloth each of* 100 *meters length is given by:*

3,5,8,9,4,2,5,9,6,4,8,12,7,5,10,10,7,5,4,10,9,5,8,13

Draw a suitable control chart and give your comments.

[B.C.A. (Kanpur) 2006, 2002; B.C.A. (Meerut) 2005]

Solution: If c_i denotes the number of defects in the i-th group, then we have the average number of defects

$$\bar{c} = \frac{3+5+8+9+\ldots+9+5+8+13}{24} = \frac{168}{24} = 7$$

∴ control limit $(\bar{c}) = 7$

3-sigma control limits for C-chart

$$\text{U.C.L.} = \bar{c} + 3\sqrt{\bar{c}} = 7 + 3\sqrt{7} = 7 + 3 \times 2.645 = 14.937$$

$$\text{L.C.L.} = \bar{c} - 3\sqrt{\bar{c}} = 7 - 3\sqrt{7} = 7 - 3 \times 2.645 = 0.937 = 0$$

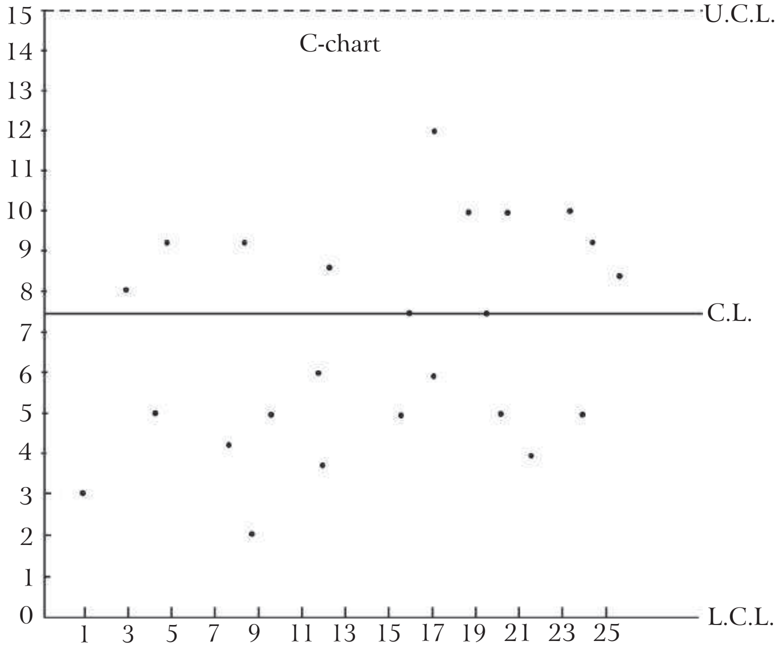


Fig. 9

Since all the points are with in the control limits, the process is in a state of statistical control.

Example 11: *Twenty five boxes, each containing 20 electric switches were randomly selected and inspected for the number of defectives in each box. The number of defectives found in each box were as follows:*

Box Number	1	2	3	4	5	6	7	8	9	10	11	12	13
No. of Defectives	3	2	1	0	4	2	1	2	3	0	2	1	2
Box Number	14	15	16	17	18	19	20	21	22	23	24	25	
No. of Defectives	0	3	5	4	2	1	3	0	3	1	2	1	

Set up control limits for a number of defectives.

[B.C.A. (Rohtak) 2011; B.C.A. (A.M.U.) 2010, 2006; B.C.A. (Agra) 2006]

Solution: Total number of defectives = $\sum_{i=1}^{k} d_i = 48$

$$\therefore \qquad \bar{p} = \frac{\sum_{i=1}^{k} d_i}{nk} = \frac{48}{20 \times 25} = 0.096$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.096 = 0.904$$

$$n\bar{p} = 20 \times 0.096 = 1.92$$

Control limits for number of defectives are:

$$\text{Control line} = n\bar{p} = 1.92$$

$$\text{U.C.L.} = n\bar{p} + 3\sqrt{n\bar{p}\bar{q}} = 1.92 + 3\sqrt{1.92 \times 0.904} = 1.92 + 3\sqrt{1.73568}$$

$$= 1.92 + 3 \times 1.31745 = 1.92 + 3.95 = 5.87$$

$$\text{L.C.L} = n\bar{p} - 3\sqrt{n\bar{p}\bar{q}} = 1.92 - 3\sqrt{1.92 \times 0.904} = 1.92 - 3\sqrt{1.73568}$$

$$= 1.92 - 3 \times 1.31745 = 1.92 - 3.95 = -2.03 = 0$$

Since the number of defectives can never be negative, the lower control limit is taken zero. The control chart for number of defectives drawn below:

Example 12: *The following are the figures of defectives in 22 lots each containing* 2,000 *rubbers belts:*

425, 430, 216, 341, 225, 322, 280, 306, 337, 305, 356, 402, 216, 264, 126, 409, 193, 326, 280, 389, 451, 420

Draw control charts for fraction defective and comment. **[B.C.A. (Meerut) 2005, 2003, 2002]**

Solution: Here $n = 2{,}000$ for each lot. If d_i and p_i are the number of defectives and sample fraction defective for ith lot respectively, then

$$p_i = \frac{d_i}{2{,}000} \qquad (i = 1, 2, 3 \ldots, 22)$$

These can be shown in following table:

S.No.	*d*	*p*	S.No.	*d*	*p*
1	425	0.2155	12	402	0.2010
2	430	0.2150	13	216	0.1080
3	216	0.1080	14	264	0.1320
4	341	0.1705	15	126	0.0630
5	225	0.1125	16	409	0.2045
6	322	0.1610	17	193	0.0965
7	280	0.1400	18	326	0.1630
8	306	0.1530	19	280	0.1400
9	337	0.1685	20	389	0.1945
10	305	0.1525	21	451	0.2255
11	356	0.1780	22	420	0.2100
					$\Sigma p_i = 3.5095$

Thus,
$$\overline{p} = \frac{\sum_{i=1}^{k} p_i}{k} = \frac{3.5095}{22} = 0.1595$$

$$\overline{q} = 1 - 0.1595 = 0.8405$$

$$\text{Control line (C.L.)} = \overline{p} = 0.1595$$

$$\text{U.C.L.} = \overline{p} + 3\sqrt{\frac{\overline{p}\overline{q}}{n}} = 0.1595 + 3\sqrt{\frac{0.1595 \times 0.8405}{2{,}000}}$$

$$= 0.1595 + 3\sqrt{0.000067} = 0.1595 + 0.0246 = 0.1841$$

$$\text{L.C.L.} = \overline{p} - 3\sqrt{\frac{\overline{p}\overline{q}}{n}} = 0.1595 - 3\sqrt{\frac{0.1595 \times 0.8405}{2{,}000}}$$

$$= 0.1595 - 3\sqrt{0.000067} = 0.1595 - 0.0246 = 0.1349$$

p-chart for fixed sample size.

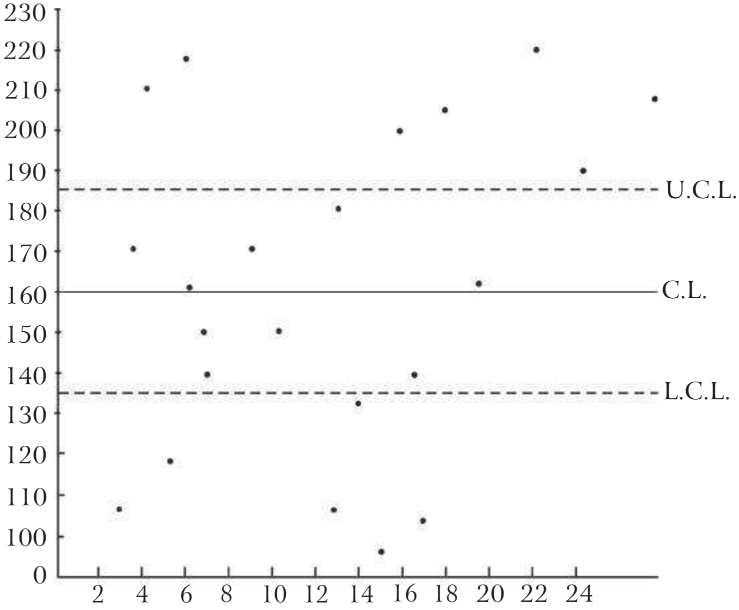


Fig. 10

From the chart it is found that a number of points fall outside the control limits, hence the process is out of control.

Example 13: *Given control limits to control the variability in production when σ is not known. Use the following values to study the variation in the production of screws when* $D_3 = 0$ *and* $D_4 = 2.282$ *for* $n = 4$.

Group No.	1	2	3	4	5	6	7	8
Range (m.m)	3	4	2	5	3	4	5	2

[B.C.A. (Kurukshetra) 2008; B.C.A. (Meerut) 2002]

Solution: We have $\overline{R} = \frac{1}{k} \sum_{i=1}^{k} R_i = \frac{28}{8} = 3.5$

Therefore, control limit for R-chart

$$\text{U.C.L.} = D_4 \overline{R} = 2.282 \times 3.5 = 7.987$$

$$\text{L.C.L.} = D_3 \overline{R} = 0 \times 3.5 = 0$$

$$\text{C.L.} = \overline{R} = 3.5$$

Example 14: *In the following data draw* $\overline{X}$*-chart and also given your comments:*

Sample Numbers	***Sample Observations***				
1	62	68	66	68	73
2	50	58	52	58	65
3	67	70	68	56	61
4	64	62	57	62	63
5	49	98	65	64	66
6	63	75	62	58	68
7	61	71	66	69	77
8	63	72	61	53	55

Given for: $n = 4, A_2 = 0.58, n = 8, A_3 = 0.373.$

[B.C.A. (Lucknow) 2009, 2007; B.C.A. (Kanpur) 2007, 2004, 2002;

B.C.A. (Agra) 2005, 2003]

Solution:

Sample Numbers	Sample Observations					Total	Mean ($\overline{X}$)	Range R
1	62	68	66	68	73	337	67.4	11
2	50	58	52	58	65	283	56.6	15
3	67	70	68	56	61	322	64.4	14
4	64	62	57	62	63	308	61.6	7
5	49	98	65	64	66	342	68.4	49
6	63	75	62	58	68	326	65.2	17
7	61	71	66	69	77	344	68.8	16
8	63	72	61	53	55	304	60.8	19
						Total	**513.2**	**148**

$$\overline{\overline{X}} = \frac{1}{k}\sum_{i=1}^{k} \overline{X}_i = \frac{513.2}{8} = 64.15$$

$$\overline{R} = \frac{1}{k}\sum_{i=1}^{k} R_i = \frac{148}{8} = 18.5$$

Control limits for $\overline{X}$-chart

$$\text{U.C.L.} = \overline{\overline{X}} + A_2\overline{R} = 64.15 + 0.58 \times 18.5 = 74.88$$

$$\text{L.C.L.} = \overline{\overline{X}} - A_2\overline{R} = 64.15 - 0.58 \times 18.5 = 53.42$$

$$\text{C.L.} = \overline{\overline{X}} - 64.45$$

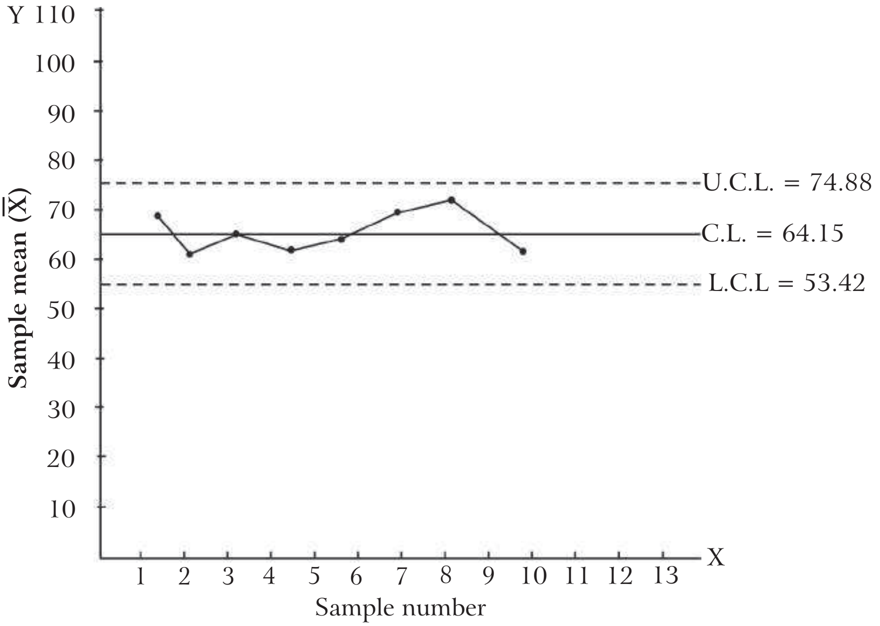


Fig. 11

Example 15: *The average number of defectives in 22-sampled lots of 2,000 rubber belts each was found to be 16%. Indicate how to construct the relevant control limits.*

[B.C.A. (Kashi Vidhyapeeth) 2011, 2009, 2007]

Solution: Number of lots $k = 22$

Number of rubber belts inspected per lot $= 2{,}000$

Average number of defectives per lot $= 16\%$

Therefore, proportion of defective per lot $= \frac{16}{100} = 0.16$

i.e. $\bar{p} = 0.16$

Control limits for *p*-chart

$$\text{U.C.L.} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.16 + 3\sqrt{\frac{0.16 \times 0.84}{2{,}000}}$$

$$= 0.16 + 3\sqrt{\frac{0.1344}{2{,}000}} = 0.16 + 3\sqrt{0.0000672}$$

$$= 0.16 + 3 \times 0.0082 = 0.16 + 0.0246 = 0.1846$$

$$\text{L.C.L.} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.16 - 3\sqrt{\frac{0.16 \times 0.84}{2{,}000}}$$

$$= 0.16 - 3\sqrt{\frac{0.1344}{2{,}000}} = 0.16 - 3\sqrt{0.0000672}$$

$$= 0.16 - 3 \times 0.0082 = 0.16 - 0.046 = 0.1354$$

Exercise

1. What do you understand by statistical quality control ? Discuss briefly the need and utility of control charts in industry.

[B.C.A. (Meerut) 2004]

2. What do you understand by statistical quality control ? Explain C-chart in detail with advantage.

[B.C.A. (Meerut) 2002]

3. Write short notes on

(i) $\overline{X}$-chart (ii) Application of C-chart

(iii) Process and product control (iv) Chance causes and assignable causes

[B.C.A. (Kanpur) 2007, 2004; B.C.A. (Meerut) 2006, 2002]

4. What are main control charts for variables ? Obtain their control limits.

[B.C.A. (Meerut) 2008, 2004]

5. What are the main control charts for attributes ? Obtain their control limits.

[B.C.A. (Lucknow) 2008; B.C.A. (A.M.U) 2006]

6. Write down the formulae for limits of C-charts.

7. Write the name of any two control charts.

8. Name the distribution.

9. What limitation of statistical quality control.

10. A company manufactures screws to a normal diameter 0.500 ± 0.030 cm. Five samples were taken randomly from the manufactured lotes and 3 measurements were taken on each sample at different lengths. Following are the readings:

S. No.	**Measurement per Sample *x* (in cm)**		
	1	**2**	**3**
1	0.488	0.489	0.505
2	0.494	0.495	0.499
3	0.498	0.515	0.487
4	0.492	0.509	0.514
5	0.490	0.508	0.499

Calculate the control limits of $\overline{X}$ and R-charts. Draw $\overline{X}$ and R-charts and examine whether the process is in statistical control. (Taking $A_2 = 1.02, D_4 = 2.57, D_3 = 0$ for $n = 3$)

11. The Number of customer complaints received daily by an organization is given below:

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Complaints	2	3	0	1	9	2	0	0	4	2	0	7	0	2	4

Does it mean that the number of complaints is under statistical control ? Establish a control scheme for the future.

12. The following data shows the value sample mean $\overline{X}$ and range R for 10 samples of size 5 each. Calculate the values for control line and control limits for $\overline{X}$-chart and R-chart and determine whether the process is under control:

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean ($\overline{X}$)	10.2	11.8	10.8	11.6	11	9.6	10.4	9.6	10.6	10
Range (R)	7	4	8	5	7	4	8	4	7	9

Assume for $n=5, A_2 = 0.577, D_3 = 0$ and $D_4 = 2.115$

Explain the construction of the control charts for variables. Give 3-σ lower and upper limits for $\overline{X}$-chart, where the sample size in n. The following number of defects were found on 20 items during an inspection schedule. calculate control limits draw a control chart and comment on your result.

3,8,5,8,6,3,4,7,8,10,2,8,6,4,5,3,0,2,6,6

14. A machine is set to deliver packets of a given weight. 10 samples of size 5 were recorded. Below are given the relevant data:

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean ($\overline{X}$)	15	17	15	18	17	14	18	15	17	16
Range (R)	7	7	4	9	8	7	12	4	11	5

Calculate the values for control line and control limits for mean chart and range chart and comment on the state of control [Conversion factors for $n=5$ are $A_2 = 0.58, D_3 = 0, D_4 = 2.11$].

15. The following shows that the number of missing rivets observed at the time of inspection of 12 air crafts. Find the control limits for the number of defects chart and comment on the state of control:

Air Craft No.	1	2	3	4	5	6	7	8	9	10	11	12
No. of Missing Rivets	1	15	13	18	10	14	13	10	20	11	22	15

16. The following data show the values of sample mean ($\overline{X}$) and range (R) for 10 samples of size 6 each. Calculate the values for control line and control limits for mean and range chart. Draw control charts and comment on the state of control:

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean ($\overline{X}$)	43	49	37	44	45	37	51	46	43	47
Range (R)	5	6	5	7	7	4	8	6	4	6

(Conversion factors for $n = 6$ are $A_2 = 0.483, D_3 = 0, D_4 = 2.004$).

17. The average percentage of defectives in 27 samples of size 15,00 each was found to be 13.7% construct p-chart for this situation. Explain how the control chart can be used to control quality

[**Hint:** $\overline{p} = 0.137$]

18. It was found that when a manufacturing process is under control, the average number of defectives per sample batch of 10 is 1.2. What would you set in a quality control chart based on the examination of defectives in sample batches of 10 ?

19. What are statistical quality control techniques ? Discuss the objectives and advantages of statistical quality control.

[B.C.A. (A.M.U.) 2010; B.C.A. (Lucknow) 2008, 2006, 2004]

20. The following are the number of defects found in 1,000 items of same industry goods inspected everyday in a month. Draw C-chart:

Days	1	2	3	4	5	6	7	8	9	10
No. of Defectives	1	1	3	7	8	1	2	6	1	1
Days	11	12	13	14	15	16	17	18	19	20
No. of Defectives	10	5	0	19	16	20	1	6	12	4
Days	21	22	23	24	25	26	27	28	29	30
No. of Defectives	5	1	8	7	9	2	3	14	6	8

21. The number of defects in 20 items each of 100 meters length is given below:

1,3,3,1,6,4,3,7,10,2,2,6,4,3,2,7,1,5,6,4

Draw the appropriate chart and give comment on your result.

22. A manufacturer finds that on an average 1 in 10 of the items produced by him is defective. A few days later he finds 20 items in a sample of 100 items defectives.

In the process out of control. Gives reasons for your answer.

(C.L. = 10, U.C.L. = 19, L.C.L. = 1 Process out of control).

23. In a capability study of a lathe used in turning a shaft to diameter of 23.75 ± 0.1 mm, a sample of 6 consecutive pieces were taken each day for 8 days and data was recorded as given below:

Days	1	2	3	4	5	6	7	8
Mean ($\overline{X}$)	23.765	23.77	23.7716	23.7767	23.7717	23.7583	23.7767	23.767
Range (R)	0.07	0.11	0.06	0.08	0.04	0.05	0.07	0.07

Plot $\overline{X}$ and R-Chart and state whether the process is in control or not. Establish the process capability and compare it with the stipulated specifications.

For a sample of size 6, take $A_2 = 0.48, D_4 = 2, D_3 = 0, d_2 = 2.534$.

Answers

10	$\text{C.L.}_{\overline{X}} = 0.4988, \text{U.C.L.}_{\overline{X}} = 0.5172, \ \text{L.C.L.}_{\overline{X}} = 0.4804,$ $\text{C.L.}_R = 0.018, \text{U.C.L.}_R = 0.0463, \ \text{L.C.L.}_R = 0,$ The process is in control.
11	$\text{C.L.}_C = 2.4, \text{U.C.L.}_C = 7.05, \text{L.C.L.}_C = 0,$ The process is not under control.
12	$\text{C.L.}_{\overline{X}} = 10.66, \text{U.C.L.}_{\overline{X}} = 14.295, \text{L.C.L.}_{\overline{X}} = 7.025,$ $\text{C.L.}_R = 0.3, \text{U.C.L.}_R = 13.32, \ \text{L.C.L.}_R = 0,$ The process is under control.
17	$\text{C.L.}_p = 0.137, \text{U.C.L.}_p = 0.164, \ \text{L.C.L.}_p = 0.110$
18	$\text{C.L.}_{np} = 1.2, \text{U.C.L.}_{np} = 2.175, \text{L.C.L.}_{np} = 0.225$

❑❑❑